



HSC Research Report

HSC/10/05

Heavy-tailed distributions in VaR calculations

Adam Misiorek*
Rafał Weron**

* Santander Consumer Bank S.A., Wrocław, Poland

** Institute of Organization and Management, Wrocław
University of Technology, Poland

Hugo Steinhaus Center
Wrocław University of Technology
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
<http://www.im.pwr.wroc.pl/~hugo/>

Heavy-tailed distributions in VaR calculations¹

Adam Misiorek² and Rafał Weron³

Abstract: The essence of the Value-at-Risk (VaR) and Expected Shortfall (ES) computations is estimation of low quantiles in the portfolio return distributions. Hence, the performance of market risk measurement methods depends on the quality of distributional assumptions on the underlying risk factors. This chapter is intended as a guide to heavy-tailed models for VaR-type calculations. We first describe stable laws and their lighter-tailed generalizations, the so-called truncated and tempered stable distributions. Next we study the class of generalized hyperbolic laws, which – like tempered stable distributions – can be classified somewhere between infinite variance stable laws and the Gaussian distribution. Then we discuss copulas, which enable us to construct a multivariate distribution function from the marginal (possibly different) distribution functions of n individual asset returns in a way that takes their dependence structure into account. This dependence structure may be no longer measured by correlation, but by other adequate functions like rank correlation, comonotonicity or tail dependence. Finally, we provide numerical examples.

Keywords: Heavy-tailed distribution; Stable distribution; Tempered stable distribution; Generalized hyperbolic distribution; Parameter estimation; Value-at-Risk (VaR); Expected Shortfall (ES); Copula; Filtered historical simulation (FHS)

JEL: C13, C16, G32

¹ Chapter prepared for the 2nd edition of *Handbook of Computational Statistics: Concepts and Methods*, J.E. Gentle, W. Härdle, Y. Mori (eds.), Springer-Verlag, *forthcoming in 2011/2012*. This version: 31.07.2010

² Santander Consumer Bank S.A., Wrocław, Poland

³ Institute of Organization and Management, Wrocław University of Technology, Poland

Heavy-tailed distributions in VaR calculations

Adam Misiorek and Rafał Weron

1 Introduction

Market risks are the prospect of financial losses – or gains – due to unexpected changes in market prices and rates. Evaluating the exposure to such risks is nowadays of primary concern to risk managers in financial and non-financial institutions alike. Since the early 1990s a commonly used market risk estimation methodology has been the Value at Risk (VaR). A VaR measure is the highest possible loss L incurred from holding the current portfolio over a certain period of time at a given confidence level (Alexander, 2008; Jorion, 2006):

$$\mathbb{P}(L > \text{VaR}) \leq 1 - c, \quad (1)$$

where c is the confidence level, typically $c \geq 95\%$. By convention, $L = -\Delta X(\tau)$, where $\Delta X(\tau)$ is the relative change (return) in portfolio value over the time horizon τ . Hence, large values of L correspond to large losses (or large negative returns).

The VaR figure has two important characteristics: (i) it provides a common consistent measure of risk across different positions and risk factors and (ii) it takes into account the correlations or dependencies between different risk factors. Because of its intuitive appeal and simplicity, it is no surprise that within a few years VaR has become the standard risk measure used around the world. However, it has a number deficiencies, among them the non-subadditivity – a sum of VaR's of two portfolios can be smaller than the VaR of the combined portfolio. To cope with these shortcomings, Artzner et al. (1999) proposed an alternative measure that satisfies the assumptions of a coherent, i.e. an adequate, risk measure. The Expected Short-

Adam Misiorek
Santander Consumer Bank S.A., Wrocław, Poland, e-mail: adam.misiorek@santanderconsumer.pl

Rafał Weron
Institute of Organization and Management, Wrocław University of Technology, 50-370 Wrocław, Poland, e-mail: rafal.weron@pwr.wroc.pl

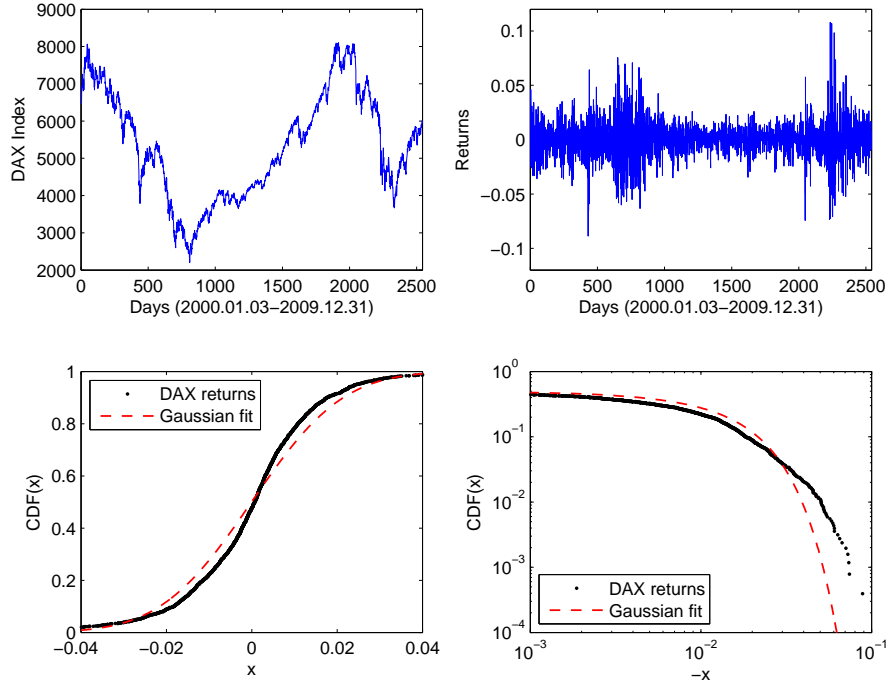


Fig. 1 *Top panels:* DAX daily closing values X_t and daily returns $\log(X_{t+1}/X_t)$ from the period January 3, 2000 – December 31, 2009. *Bottom panels:* Gaussian fit to the DAX daily returns empirical cumulative distribution function (CDF). For better exposition of the fit in the central part of the distribution the range is limited to $\pm 4\%$. The right panel is a magnification of the left tail fit on a double logarithmic scale clearly showing the discrepancy between the data and the normal distribution.

fall (ES), also called Expected Tail Loss or Conditional VaR, is the expected value of the losses in excess of VaR, i.e. $ES = \mathbb{E}(L|L > VaR)$. It is interesting to note, that the notion of Expected Shortfall has been familiar to insurance practitioners for decades. It is very similar to the mean excess function which is used to characterize claim size distributions (see Chapter 13 in Cizek, Härdle and Weron, 2005).

The essence of the VaR and ES computations is estimation of low quantiles in the portfolio return distributions. Hence, the performance of market risk measurement methods depends on the quality of distributional assumptions on the underlying risk factors. Many of the concepts in theoretical and empirical finance developed over the past decades – including the classical portfolio theory, the Black-Scholes-Merton option pricing model and even the RiskMetrics variance-covariance approach to VaR – rest upon the assumption that asset returns follow a normal distribution. But is this assumption justified by empirical data?

No, it is not! It has been long known that asset returns are not normally distributed. Rather, the empirical observations exhibit excess kurtosis (fat tails). In Figure 1 we plot a ten-year history (January 3, 2000 – December 31, 2009) of

the Deutsche Aktienindex (DAX) index, its returns (or log-returns) and the distribution of the returns. The contrast with the Gaussian law is striking. This heavy tailed or leptokurtic character of the distribution of price changes has been repeatedly observed in various markets and may be quantitatively measured by the kurtosis in excess of 3, a value obtained for the normal distribution (Guillaume et al., 1997; Rachev and Mittnik, 2000).

Interestingly, the problem of the underestimation of risk by the Gaussian distribution has been dealt with by the regulators in an ad hoc way. The Basle Committee on Banking Supervision (1995) suggested that for the purpose of determining minimum capital reserves financial institutions use a ten day VaR at the $c = 99\%$ confidence level multiplied by a safety factor $s \in [3, 4]$, with the exact value of s depending on the past performance of the model. It has been argued by Stahl (1997) and Danielsson, Hartmann and De Vries (1998) that the range of the safety factor comes from the heavy-tailed nature of the returns distribution. Indeed, if we assume that the asset returns distribution is symmetric and has finite variance σ^2 then from Chebyshev's inequality we obtain $\mathbb{P}(L \geq \varepsilon) \leq \sigma^2/2\varepsilon^2$, where L represents the random loss over the specified time horizon. So if we want to calculate the upper bound for a 99% VaR, setting $\sigma^2/2\varepsilon^2 = 1\%$ yields $\varepsilon = 7.07\sigma$, which in turn implies that $\text{VaR}_{99\%} \leq 7.07\sigma$. However, if we assumed a Gaussian distribution of returns then we would have $\text{VaR}_{99\%} \leq 2.33\sigma$, which is roughly three times lower than the bound obtained for a heavy-tailed, finite variance distribution.

Having said this much about the inadequacy of the Gaussian distribution for financial modeling and risk management we have no other choice but offer some heavy-tailed alternatives. We have to mention, though, that all distributional classes described in this chapter present computational challenge. Large parts of the text are thus devoted to numerical issues. In Section 2 we deal with the historically earliest alternative – the stable laws – and briefly characterize their recent generalizations – the so-called truncated and tempered stable distributions. Further, in Section 3 we study the class of generalized hyperbolic laws. Finally, in Section 4 we introduce the notion of copulas and discuss the relation between VaR, asset portfolios and heavy tails.

2 Stable distributions

2.1 Definitions and basic properties

Since the pioneering work of Louis Bachelier in 1900, financial asset returns have been modeled by the Gaussian distribution. The theoretical rationale for this comes from the Central Limit Theorem (CLT), which states that the sum of a large number of independent, identically distributed (i.i.d.) variables – say, decisions of investors – from a finite-variance distribution will be (asymptotically) normally distributed. However, empirical evidence indicates that financial asset returns tend to have heav-

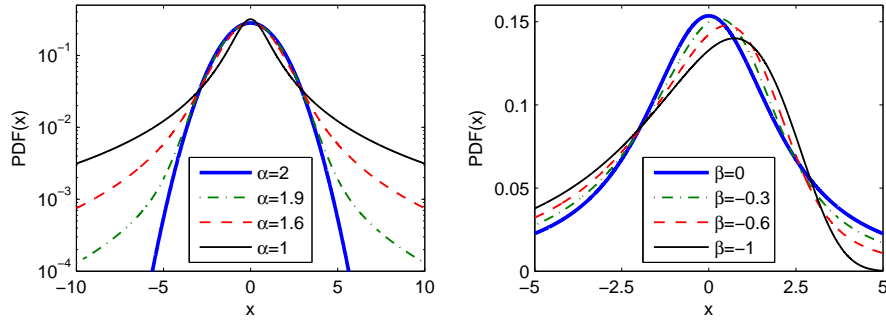


Fig. 2 *Left panel:* A semilog plot of symmetric ($\beta = \mu = 0$) stable probability density functions (PDFs) for four different values of α showing the dependence on the tail exponent. The Gaussian ($\alpha = 2$) PDF forms a parabola and is the only stable density with exponential tails. *Right panel:* A plot of stable PDFs for $\alpha = 1.1$ and four different values of β showing the dependence on the skewness parameter.

ier tails than Gaussian. Possible reasons for the failure of the CLT are infinite-variance distributions of the variables, non-identical distributions of the variables, dependences between the variables or any combination of the three.

The dependence issue is hard to tackle analytically, however, if only the finite variance assumption is dropped we have a readily usable solution. Namely, the generalized version of the CLT states that the limiting distribution of sums of such variables is stable (Nolan, 2010). This, together with the fact that stable distributions are leptokurtic and can accommodate fat tails and asymmetry, provides us with a theoretically justified modeling tool. Indeed, as early as in the 1960s stable laws were proposed as an alternative model for asset returns (Mandelbrot, 1963).

Stable laws – also called α -stable, stable Paretian or Lévy stable – were introduced by Lévy (1925) during his investigations of the behavior of sums of independent random variables. The name ‘stable’ reflects the fact that a sum of two independent random variables having a stable distribution with the same index α is again stable with index α . Recall, that this invariance property holds also for Gaussian variables. In fact, the Gaussian distribution is stable with $\alpha = 2$.

The stable distribution requires four parameters for complete description. The index of stability $\alpha \in (0, 2]$, also called the tail index, tail exponent or characteristic exponent, determines the rate at which the tails of the distribution taper off, see the left panel in Figure 2. The skewness parameter $\beta \in [-1, 1]$ defines the asymmetry. When $\beta > 0$, the distribution is skewed to the right, i.e. the right tail is thicker, see the right panel in Figure 2. When it is negative, it is skewed to the left. When $\beta = 0$, the distribution is symmetric about the mode (the peak) of the distribution. As α approaches 2, β loses its effect and the distribution approaches the Gaussian distribution regardless of β . The last two parameters, $\sigma > 0$ and $\mu \in \mathbb{R}$, are the usual scale and location parameters, respectively.

From a practitioner’s point of view the crucial drawback of the stable distribution is that, with the exception of three special cases, its probability density function

tion (PDF) and cumulative distribution function (CDF) do not have closed form expressions. These exceptions include the well known Gaussian ($\alpha = 2$) law, whose density is given by:

$$f_G(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad (2)$$

and the lesser known Cauchy ($\alpha = 1, \beta = 0$) and Lévy ($\alpha = 0.5, \beta = 1$) laws.

Hence, the stable distribution can be most conveniently described by its characteristic function (CF) – the inverse Fourier transform of the PDF. The most popular parameterization of the characteristic function $\phi(t)$ of $X \sim S_\alpha(\sigma, \beta, \mu)$, i.e. a stable random variable with parameters α, σ, β and μ , is given by (Samorodnitsky and Taqqu, 1994; Weron, 1996):

$$\log \phi(t) = \begin{cases} -\sigma^\alpha |t|^\alpha \{1 - i\beta \text{sign}(t) \tan \frac{\pi\alpha}{2}\} + i\mu t, & \alpha \neq 1, \\ -\sigma |t| \{1 + i\beta \text{sign}(t) \frac{2}{\pi} \log |t|\} + i\mu t, & \alpha = 1. \end{cases} \quad (3)$$

Note, that the traditional scale parameter σ of the Gaussian distribution is not the same as σ in the above representation. A comparison of formulas (2) and (3) yields the relation: $\sigma_{\text{Gaussian}} = \sqrt{2}\sigma$.

For numerical purposes, it is often useful to use Nolan's (1997) parameterization:

$$\log \phi_0(t) = \begin{cases} -\sigma^\alpha |t|^\alpha \{1 + i\beta \text{sign}(t) \tan \frac{\pi\alpha}{2} [(\sigma|t|)^{1-\alpha} - 1]\} + i\mu_0 t, & \alpha \neq 1, \\ -\sigma |t| \{1 + i\beta \text{sign}(t) \frac{2}{\pi} \log(\sigma|t|)\} + i\mu_0 t, & \alpha = 1, \end{cases} \quad (4)$$

which yields a CF (and hence the PDF and CDF) jointly continuous in all four parameters. The location parameters of the two representations (S and S^0) are related by $\mu = \mu_0 - \beta\sigma \tan \frac{\pi\alpha}{2}$ for $\alpha \neq 1$ and $\mu = \mu_0 - \beta\sigma \frac{2}{\pi} \log \sigma$ for $\alpha = 1$. Moreover, when $\alpha > 1$ the mean of the distribution exists and is equal to μ .

The latter is a result of a more general property: the p th moment of a stable random variable is finite if and only if $p < \alpha$. Hence, when $\alpha < 2$ the variance is infinite and the tails exhibit a power-law behavior (i.e. they are asymptotically equivalent to a Pareto law). More precisely, using a CLT type argument it can be shown that (Janicki and Weron, 1994a; Samorodnitsky and Taqqu, 1994):

$$\begin{cases} \lim_{x \rightarrow \infty} x^\alpha \mathbb{P}(X > x) = C_\alpha(1 + \beta)\sigma^\alpha, \\ \lim_{x \rightarrow \infty} x^\alpha \mathbb{P}(X < -x) = C_\alpha(1 - \beta)\sigma^\alpha, \end{cases} \quad (5)$$

where $C_\alpha = (2 \int_0^\infty x^{-\alpha} \sin(x) dx)^{-1} = \frac{1}{\pi} \Gamma(\alpha) \sin \frac{\pi\alpha}{2}$. The convergence to the power-law tail varies for different α 's and is slower for larger values of the tail index. Moreover, the tails of stable CDFs exhibit a crossover from an approximate power decay with exponent $\alpha > 2$ to the true tail with exponent α . This phenomenon is more visible for large α 's (Weron, 2001).

2.2 Truncating or tempering the tails

Mandelbrot's (1963) seminal work on applying stable distributions in finance gained support in the first few years after its publication, but subsequent works have questioned the stable distribution hypothesis, in particular, the stability under summation (for a review see Rachev and Mittnik, 2000). Over the next few years, the stable law temporarily lost favor and alternative processes were suggested as mechanisms generating stock returns. In the mid 1990s the stable distribution hypothesis has made a dramatic comeback, at first in the econophysics literature. Several authors have found a very good agreement of high-frequency returns with a stable distribution up to six standard deviations away from the mean (Cont, Potters and Bouchaud, 1997). For more extreme observations, however, the distribution they found fell off approximately exponentially. To cope with such observations the so called truncated Lévy distributions (TLD) were introduced by Mantegna and Stanley (1994). The original definition postulated a sharp truncation of the stable PDF at some arbitrary point. Later, however, exponential smoothing was proposed by Koponen (1995) leading to the following characteristic function:

$$\log \phi(t) = -\frac{\sigma^\alpha}{\cos \frac{\pi\alpha}{2}} \left[(t^2 + \lambda^2)^{\alpha/2} \cos \left\{ \alpha \arctan \frac{|t|}{\lambda} \right\} - \lambda^\alpha \right], \quad (6)$$

where $\alpha \neq 1$ is the tail exponent, σ is the scale parameter and λ is the truncation coefficient (for simplicity β and μ are set to zero here). Clearly the symmetric TLD reduces to the symmetric stable distribution ($\beta = \mu = 0$) when $\lambda = 0$. For small and intermediate returns the TLD behaves like a stable distribution, but for extreme returns the truncation causes the distribution to converge to a Gaussian (hence, all moments are finite), see Figure 3. Thus the observation that the asset returns distribution is a TLD explains both the short-term stable behavior and the long run convergence to the normal distribution (for interesting insights on the CLT-type behavior of the TLD see a recent paper of Grabchak and Samorodnitsky, 2010).

The (exponentially smoothed) TLD was not recognized in finance until the introduction of the KoBoL (Boyarchenko and Levendorskii, 2000) and CGMY models (Carr et al., 2002). Around this time Rosinski coined the term under which the TLD is known today in the mathematics literature – tempered stable distribution (TSD; see Rosinski, 2007).

Despite the interesting statistical properties, the TSDs (TLDs) have not been applied extensively to date. The most probable reason for this being the complicated definition of the TSD. Like for stable distributions, only the characteristic function is known. No closed form formulas exist for the density or the distribution functions. No integral formulas, like Zolotarev's (1986) for the stable laws (see Section 2.3), have been discovered to date. Hence, statistical inference is, in general, limited to ML utilizing the FFT technique for approximating the PDF (Bianchi et al., 2010; Grabchak, 2008). Moreover, compared to the stable distribution, the TSD introduces one more parameter making the estimation procedure even more complicated. Other parameter fitting techniques proposed so far com-

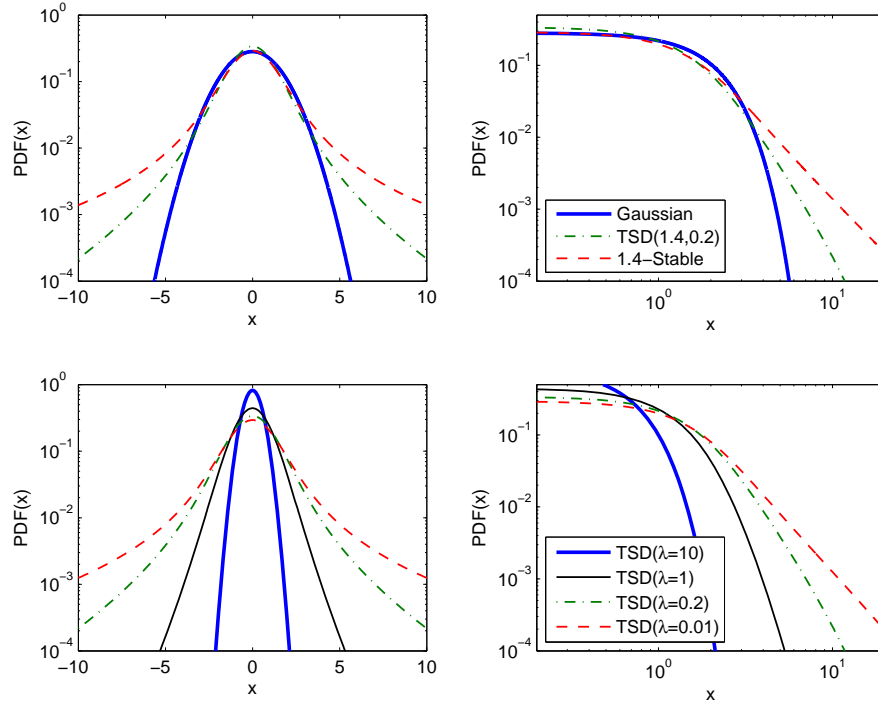


Fig. 3 *Top panels:* Semilog and loglog plots of symmetric 1.4-stable, symmetric tempered stable with $\alpha = 1.4$ and $\lambda = 0.2$, and Gaussian PDFs. *Bottom panels:* Semilog and loglog plots of symmetric tempered stable PDFs with $\alpha = 1.4$ and four different truncation coefficients: $\lambda = 10, 1, 0.2, 0.01$. Note, that for large λ 's the distribution approaches the Gaussian (though with a different scale) and for small λ 's the stable law with the same shape parameter α .

prise a combination of ad hoc approaches and moment matching (Boyarchenko and Levendorskii, 2000; Matacz, 2000). Apart from a few special cases, also the simulation of TSD variables is cumbersome and numerically demanding (Bianchi et al., 2010; Kawai and Masuda, 2010; Poirot and Tankov, 2006).

2.3 Computation of stable density and distribution functions

The lack of closed form formulas for most stable densities and distribution functions has negative consequences. Numerical approximation or direct numerical integration have to be used, leading to a drastic increase in computational time and loss of accuracy. Of all the attempts to be found in the literature a few are worth mentioning. DuMouchel (1971) developed a procedure for approximating the stable CDF using Bergström's series expansion. Depending on the particular range of α and β , Holt and Crow (1973) combined four alternative approximations to compute the stable

PDF. Both algorithms are computationally intensive and time consuming, making maximum likelihood estimation a nontrivial task, even for modern computers. In the late 1990s, two more efficient techniques have been proposed.

Mittnik, Doganoglu and Chenyao (1999) exploited the PDF–CF relationship and applied the fast Fourier transform (FFT). However, for data points falling between the equally spaced FFT grid nodes an interpolation technique has to be used. The authors suggested that linear interpolation suffices in most practical applications, see also Rachev and Mittnik (2000). Taking a larger number of grid points increases accuracy, however, at the expense of higher computational burden. Setting the number of grid points to $N = 2^{13}$ and the grid spacing to $h = 0.01$ allows to achieve comparable accuracy to the direct integration method (see below), at least for a range of α 's typically found for financial data ($1.6 < \alpha < 1.9$).

As for the computational speed, the FFT based approach is faster for large samples, whereas the direct integration method favors small data sets since it can be computed at any arbitrarily chosen point. Mittnik, Doganoglu and Chenyao (1999) report that for $N = 2^{13}$ the FFT based method is faster for samples exceeding 100 observations and slower for smaller data sets. We must stress, however, that the FFT based approach is not as universal as the direct integration method – it is efficient only for large alpha's and only as far as the PDF calculations are concerned. When computing the CDF the former method must numerically integrate the density, whereas the latter takes the same amount of time in both cases.

The direct integration method, proposed by Nolan (1997, 1999), consists of a numerical integration of Zolotarev's (1986) formulas for the density or the distribution function. To save space we state only the formulas for the PDF. Complete formulas can be also found in Cizek, Härdle and Weron (2005), Chapter 1.

Set $\zeta = -\beta \tan \frac{\pi\alpha}{2}$. Then the density $f(x; \alpha, \beta)$ of a standard stable random variable in representation S^0 , i.e. $X \sim S_\alpha^0(1, \beta, 0)$, can be expressed as (note, that Zolotarev (1986, Section 2.2) used another parametrization):

- when $\alpha \neq 1$ and $x \neq \zeta$:

$$f(x; \alpha, \beta) = \frac{\alpha(x - \zeta)^{\frac{1}{\alpha}-1}}{\pi |\alpha - 1|} \int_{-\theta_0}^{\frac{\pi}{2}} V(\theta; \alpha, \beta) \exp \left\{ -(x - \zeta)^{\frac{\alpha}{\alpha-1}} V(\theta; \alpha, \beta) \right\} d\theta,$$

for $x > \zeta$ and $f(x; \alpha, \beta) = f(-x; \alpha, -\beta)$ for $x < \zeta$,

- when $\alpha \neq 1$ and $x = \zeta$:

$$f(x; \alpha, \beta) = \frac{\Gamma(1 + \frac{1}{\alpha}) \cos(\frac{\pi}{2\alpha})}{\pi(1 + \zeta^2)^{\frac{1}{2\alpha}}},$$

- when $\alpha = 1$:

$$f(x; 1, \beta) = \begin{cases} \frac{1}{2|\beta|} e^{\frac{\pi x}{2\beta}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} V(\theta; 1, \beta) \exp \left\{ -e^{\frac{\pi x}{2\beta}} V(\theta; 1, \beta) \right\} d\theta, & \beta \neq 0, \\ \frac{1}{\pi(1+x^2)}, & \beta = 0, \end{cases}$$

where

$$\xi = \begin{cases} \frac{1}{\alpha} \arctan(-\zeta), & \alpha \neq 1, \\ \frac{\pi}{2}, & \alpha = 1, \end{cases} \quad (7)$$

and

$$V(\theta; \alpha, \beta) = \begin{cases} (\cos \alpha \xi)^{\frac{1}{\alpha-1}} \left(\frac{\cos \theta}{\sin \alpha(\xi+\theta)} \right)^{\frac{\alpha}{\alpha-1}} \frac{\cos\{\alpha\xi+(\alpha-1)\theta\}}{\cos \theta}, & \alpha \neq 1, \\ \frac{2}{\pi} \left(\frac{\frac{\pi}{2} + \beta\theta}{\cos \theta} \right) \exp \left\{ \frac{1}{\beta} \left(\frac{\pi}{2} + \beta\theta \right) \tan \theta \right\}, & \alpha = 1, \beta \neq 0. \end{cases}$$

To the best of our knowledge, currently no statistical computing environment offers the computation of stable density and distribution functions in its standard release. Users have to rely on third-party libraries or commercial products. A few are worth mentioning. The standalone program STABLE (downloadable from John Nolan's web page: academic2.american.edu/~jpnolan/stable/stable.html) is probably the most efficient. It was written in Fortran and calls several external IMSL routines, see Nolan (1997) for details. Apart from speed, the STABLE program also exhibits high relative accuracy (ca. 10^{-13} ; for default tolerance settings) for extreme tail events and 10^{-10} for values used in typical financial applications (like approximating asset return distributions). The STABLE program is also available in library form through Robust Analysis Inc. (www.robustanalysis.com). This library provides interfaces to Matlab, S-plus/R and Mathematica.

In the late 1990s Diethelm Würtz has initiated the development of Rmetrics, an open source collection of S-plus/R software packages for computational finance (www.rmetrics.org). In the *fBasics* package stable PDF and CDF calculations are performed using the direct integration method, with the integrals being computed by R's function *integrate*.

The FFT based approach is utilized in Cognity, a commercial risk management platform that offers derivatives pricing and portfolio optimization based on the assumption of stably distributed returns (www.finanalytica.com). The FFT implementation is also available in Matlab (*stablepdf_fft.m*) from the Statistical Software Components repository (ideas.repec.org/c/boc/bocode/m429004.html).

2.4 Simulation of stable variables

Simulating sequences of stable random variables is not straightforward, since there are no analytic expressions for the inverse $F^{-1}(x)$ nor the CDF $F(x)$ itself. All standard approaches like the rejection or the inversion methods would require tedious computations. See Chapter ?? for a review of non-uniform random number generation techniques.

A much more elegant and efficient solution was proposed by Chambers, Mallows and Stuck (1976). They noticed that a certain integral formula derived by Zolotarev (1964) yielded the following algorithm:

- generate a random variable U uniformly distributed on $(-\frac{\pi}{2}, \frac{\pi}{2})$ and an independent exponential random variable W with mean 1;
- for $\alpha \neq 1$ compute:

$$X = (1 + \xi^2)^{\frac{1}{2\alpha}} \frac{\sin\{\alpha(U + \xi)\}}{\{\cos(U)\}^{1/\alpha}} \left[\frac{\cos\{U - \alpha(U + \xi)\}}{W} \right]^{\frac{1-\alpha}{\alpha}}, \quad (8)$$

- for $\alpha = 1$ compute:

$$X = \frac{1}{\xi} \left\{ \left(\frac{\pi}{2} + \beta U \right) \tan U - \beta \log \left(\frac{\frac{\pi}{2} W \cos U}{\frac{\pi}{2} + \beta U} \right) \right\}, \quad (9)$$

where ξ is given by eqn. (7). This algorithm yields a random variable $X \sim S_\alpha(1, \beta, 0)$, in representation (3). For a detailed proof see Weron (1996).

Given the formulas for simulation of a standard stable random variable, we can easily simulate a stable random variable for all admissible values of the parameters α , σ , β and μ using the following property. If $X \sim S_\alpha(1, \beta, 0)$ then

$$Y = \begin{cases} \sigma X + \mu, & \alpha \neq 1, \\ \sigma X + \frac{2}{\pi} \beta \sigma \log \sigma + \mu, & \alpha = 1, \end{cases} \quad (10)$$

is $S_\alpha(\sigma, \beta, \mu)$. It is interesting to note that for $\alpha = 2$ (and $\beta = 0$) the Chambers-Mallows-Stuck (CMS) method reduces to the well known Box-Muller algorithm for generating Gaussian random variables (Janicki and Weron, 1994b).

Many other approaches have been proposed in the literature, including application of Bergström and LePage series expansions, see Mantegna (1994) and Janicki and Kokoszka (1992), respectively. However, the CMS method is regarded as the fastest and the most accurate. On a PC equipped with a Core 2 Duo 2.66 GHz CPU one million variables are generated in about 1.03 seconds (using *stablernd.m* from the SSC repository: ideas.repec.org/c/boc/bocode/m429003.html), compared to about 0.058 seconds for one million standard normal random variables obtained via the Box-Muller algorithm (*randn.m* in Matlab). Because of its unquestioned superiority and relative simplicity, the CMS method is implemented in some statistical computing environments (e.g. the *rstable* function in S-plus/R) even if no other routines related to stable distributions are provided.

2.5 Estimation of parameters

The lack of known closed-form density functions for all but a few members of the stable family results in a considerable numerical complexity of statistical inference for these distributions. For instance, maximum likelihood (ML) estimates have to be based on numerical approximations or direct numerical integration of the formulas presented in Section 2.3. Consequently, ML estimation is difficult to implement and

time consuming for samples encountered in modern finance. However, there are also other numerical methods that have been found useful in practice and are discussed in this section.

Given a sample x_1, \dots, x_n of i.i.d. $S_\alpha(\sigma, \beta, \mu)$ observations, in what follows, we provide estimates $\hat{\alpha}$, $\hat{\sigma}$, $\hat{\beta}$ and $\hat{\mu}$ of all four stable law parameters. We start the discussion with the simplest, fastest and ... least accurate quantile methods, then develop the slower, yet much more accurate sample CF methods and, finally, conclude with the slowest but most accurate ML approach.

All presented methods work quite well assuming that the sample under consideration is indeed stable. However, testing for stability is not an easy task. Despite some more or less successful attempts (Brcich, Iskander and Zoubir, 2005; Cizek, Härdle and Weron, 2005; Paoletta, 2001; Matsui and Takemura, 2008), there are no standard, widely-accepted tests for assessing stability. A possible remedy may be to use one of the computationally less demanding tail exponent estimators (Fan, 2006; Mitnik and Paoletta, 1999) or simply ‘visual inspection’ to see whether the empirical densities resemble those of stable laws (Nolan, 2001; Weron, 2001).

2.5.1 Sample quantile methods

The origins of sample quantile methods for stable laws go back to Fama and Roll (1971), who provided very simple estimates for parameters of symmetric ($\beta = 0, \mu = 0$) stable laws with $\alpha > 1$. A decade later McCulloch (1986) generalized their method and provided consistent estimators of all four stable parameters (with the restriction $\alpha \geq 0.6$). After McCulloch define:

$$v_\alpha = \frac{x_{0.95} - x_{0.05}}{x_{0.75} - x_{0.25}} \quad \text{and} \quad v_\beta = \frac{x_{0.95} + x_{0.05} - 2x_{0.50}}{x_{0.95} - x_{0.05}}, \quad (11)$$

where x_f denotes the f -th population quantile, so that $S_\alpha(\sigma, \beta, \mu)(x_f) = f$. Statistics v_α and v_β are functions of α and β only, i.e. they are independent of both σ and μ . This relationship may be inverted and the parameters α and β may be viewed as functions of v_α and v_β . Substituting v_α and v_β by their sample values and applying linear interpolation between values found in tables given in McCulloch (1986) yields estimators $\hat{\alpha}$ and $\hat{\beta}$. Scale and location parameters, σ and μ , can be estimated in a similar way. However, due to the discontinuity of the CF for $\alpha = 1$ and $\beta \neq 0$ in representation (3), this procedure is much more complicated.

In a recent paper, Dominicy and Veredas (2010) further extended the quantile approach by introducing the method of simulated quantiles. It is a promising approach which can also handle multidimensional cases as, for instance, the joint estimation of N univariate stable distributions (but with the constraint of a common tail index).

2.5.2 Sample characteristic function methods

Given an i.i.d. random sample x_1, \dots, x_n of size n , define the sample CF by: $\hat{\phi}(t) = \frac{1}{n} \sum_{j=1}^n \exp(itx_j)$. Since $|\hat{\phi}(t)|$ is bounded by unity all moments of $\hat{\phi}(t)$ are finite and, for any fixed t , it is the sample average of i.i.d. random variables $\exp(itx_j)$. Hence, by the law of large numbers, $\hat{\phi}(t)$ is a consistent estimator of the CF $\phi(t)$.

To the best of our knowledge, Press (1972) was the first to use the sample CF in the context of statistical inference for stable laws. He proposed a simple estimation method for all four parameters, called the method of moments, based on transformations of the CF. However, the convergence of this method to the population values depends on the choice of four estimation points, whose selection is problematic.

Koutrouvelis (1980) presented a much more accurate regression-type method which starts with an initial estimate of the parameters and proceeds iteratively until some prespecified convergence criterion is satisfied. Each iteration consists of two weighted regression runs. The number of points to be used in these regressions depends on the sample size and starting values of α . Typically no more than two or three iterations are needed. The speed of the convergence, however, depends on the initial estimates and the convergence criterion.

The regression method is based on the following observations concerning the CF $\phi(t)$. First, from (3) we can easily derive:

$$\log(-\log |\phi(t)|^2) = \log(2\sigma^\alpha) + \alpha \log |t|. \quad (12)$$

The real and imaginary parts of $\phi(t)$ are for $\alpha \neq 1$ given by:

$$\Re\{\phi(t)\} = \exp(-|\sigma t|^\alpha) \cos \left[\mu t + |\sigma t|^\alpha \beta \operatorname{sign}(t) \tan \frac{\pi\alpha}{2} \right], \quad (13)$$

$$\Im\{\phi(t)\} = \exp(-|\sigma t|^\alpha) \sin \left[\mu t + |\sigma t|^\alpha \beta \operatorname{sign}(t) \tan \frac{\pi\alpha}{2} \right]. \quad (14)$$

Apart from considerations of principal values, equations (13)-(14) lead to:

$$\arctan \left(\frac{\Im\{\phi(t)\}}{\Re\{\phi(t)\}} \right) = \mu t + \beta \sigma^\alpha \tan \frac{\pi\alpha}{2} \operatorname{sign}(t) |t|^\alpha. \quad (15)$$

Equation (12) depends only on α and σ and suggests that we can estimate these two parameters by regressing $y = \log(-\log |\phi_n(t)|^2)$ on $w = \log |t|$ in the model: $y_k = m + \alpha w_k + \varepsilon_k$, where t_k is an appropriate set of real numbers, $m = \log(2\sigma^\alpha)$, and ε_k denotes an error term. Koutrouvelis (1980) proposed to use $t_k = \frac{\pi k}{25}$, $k = 1, 2, \dots, K$; with K ranging between 9 and 134 for different values of α and sample sizes.

Once $\hat{\alpha}$ and $\hat{\sigma}$ have been obtained and α and σ have been fixed at these values, estimates of β and μ can be obtained using (15). Next, the regressions are repeated with $\hat{\alpha}$, $\hat{\sigma}$, $\hat{\beta}$ and $\hat{\mu}$ as the initial parameters. The iterations continue until a pre-specified convergence criterion is satisfied. Koutrouvelis proposed to use Fama and Roll's (1971) formula and the 25% truncated mean for initial estimates of σ and μ , respectively.

Table 1 Comparison of McCulloch's quantile technique, the regression approach of Koutrouvelis and the method of Kogon and Williams for 1000 simulated samples of two thousand $S_{1.7}(0.005, 0.1, 0.001)$ random numbers each. Parameter estimates are mean values over 1000 samples. Values of the Mean Absolute Percentage Error ($MAPE_{\theta} = \frac{1}{n} \sum_{i=1}^n |\hat{\theta} - \theta|/\theta$) are given in parentheses. In the last column CPU time factors (average computational times relative to the method of Kogon and Williams) for one sample of 2000 random variables are provided.

Method	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\beta}$	$\hat{\mu}$	CPU time factor
McCulloch	1.7070 (2.72%)	0.0050 (2.14%)	0.1108 (108.97%)	0.0013 (29.90%)	0.33×
Koutrouvelis	1.7021 (1.66%)	0.0050 (1.63%)	0.0933 (91.99%)	0.0012 (27.76%)	5.62×
Kogon-Williams	1.7030 (1.91%)	0.0050 (1.71%)	0.0934 (99.64%)	0.0010 (16.72%)	1.00×

Kogon and Williams (1998) eliminated this iteration procedure and simplified the regression method. For initial estimation they applied McCulloch's method, worked with the continuous representation (4) of the CF instead of the classical one (3) and used a fixed set of only 10 equally spaced frequency points t_k . In terms of computational speed their method compares favorably to the original method of Koutrouvelis, see Table 1. It has a significantly better performance near $\alpha = 1$ and $\beta \neq 0$ due to the elimination of discontinuity of the CF. However, it returns slightly worse results for other values of α . Matlab implementations of McCulloch's quantile technique (*stabcull.m*) and the regression approach of Koutrouvelis (*stabreg.m*) are distributed with the MFE Toolbox accompanying the monograph of Weron (2006) and can be downloaded from www.ioz.pwr.wroc.pl/pracownicy/weron/MFE.htm.

A typical performance of the described estimators is summarized in Table 1. McCulloch's quantile technique, the regression approach of Koutrouvelis and the method of Kogon and Williams were applied to 1000 simulated samples of two thousand $S_{1.7}(0.005, 0.1, 0.001)$ random numbers each. McCulloch's method yielded the worst, but acceptable estimates and computational time significantly lower than the regression approaches. On the other hand, both the Koutrouvelis and the Kogon-Williams implementations yielded good estimators with the latter performing considerably faster, but slightly less accurate. We have to say, though, that all methods had problems with estimating β . Like it or not, our search for the optimal estimation technique is not over yet. We have no other choice but turn to the last resort – the ML method.

2.5.3 Maximum likelihood method

The maximum likelihood (ML) estimation scheme for stable distributions does not differ from that for other laws, at least as far as the theory is concerned. For a vector of observations $x = (x_1, \dots, x_n)$, the ML estimate of the parameter vector $\theta = (\alpha, \sigma, \beta, \mu)$ is obtained by maximizing the log-likelihood function:

$$L_{\theta}(x) = \sum_{i=1}^n \log \tilde{f}(x_i; \theta), \quad (16)$$

where $\tilde{f}(\cdot; \theta)$ is the stable density function. The tilde denotes the fact that, in general, we do not know the explicit form of the stable PDF and have to approximate it numerically. The ML methods proposed in the literature differ in the choice of the approximating algorithm. However, all of them have an appealing common feature – under certain regularity conditions the ML estimator is asymptotically normal with the variance specified by the Fischer information matrix (DuMouchel, 1973). The latter can be approximated either by using the Hessian matrix arising in maximization or, as in Nolan (2001), by numerical integration.

Because of computational complexity there are only a few documented attempts of estimating stable law parameters via maximum likelihood worth mentioning. DuMouchel (1971) developed an approximate ML method, which was based on grouping the data set into bins and using a combination of means to compute the density (FFT for the central values of x and series expansions for the tails) to compute an approximate log-likelihood function. This function was then numerically maximized.

Much better, in terms of accuracy and computational time, are more recent ML estimation techniques. Mitnik et al. (1999) utilized the FFT approach for approximating the stable density function, whereas Nolan (2001) used the direct integration method. Both approaches are comparable in terms of efficiency. The differences in performance are the result of different approximation algorithms, see Section 2.3. Matsui and Takemura (2006) further improved Nolan’s method for the boundary cases, i.e. in the tail and mode of the densities and in the neighborhood of the Cauchy and the Gaussian distributions, but only in the symmetric stable case.

As Ojeda (2001) observes, the ML estimates are almost always the most accurate, closely followed by the regression-type estimates and McCulloch’s quantile method. However, ML estimation techniques are certainly the slowest of all the discussed methods. For instance, ML estimation for a sample of 2000 observations using a gradient search routine which utilizes the direct integration method is over 11 thousand (!) times slower than the Kogon-Williams algorithm (calculations performed on a PC running STABLE ver. 3.13; see Section 2.3 where the program was briefly described). Clearly, the higher accuracy does not justify the application of ML estimation in many real life problems, especially when calculations are to be performed on-line. For this reason the program STABLE offers an alternative – a fast quasi ML technique. It quickly approximates stable densities using a 3-dimensional spline interpolation based on pre-computed values of the standardized stable density on a grid of (x, α, β) values. At the cost of a large array of coefficients, the interpolation is highly accurate over most values of the parameter space and relatively fast – only ca. 13 times slower than the Kogon-Williams algorithm for a sample of 2000 observations.

2.5.4 Alternative methods

Besides the popular methods discussed so far other estimation algorithms have been proposed in the literature. A Bayesian Markov chain Monte Carlo (MCMC) approach was initiated by Buckle (1995). It was later modified by Lombardi (2007) who used an approximated version of the likelihood, instead of the twice slower Gibbs sampler, and by Peters, Sisson and Fan (2009) who proposed likelihood-free Bayesian inference for stable models.

In a recent paper Garcia, Renault and Veredas (2010) estimate the stable law parameters with (constrained) indirect inference, a method particularly suited to situations where the model of interest is difficult to estimate but relatively easy to simulate. They use the skewed- t distribution as an auxiliary model, since it has the same number of parameters as the stable with each parameter playing a similar role.

3 Generalized hyperbolic distributions

3.1 Definitions and basic properties

The hyperbolic law saw its appearance in finance in the mid-1990s, when a number of authors reported that it provides a very good model for the distributions of daily stock returns from a number of leading German enterprises (Eberlein and Keller, 1995; Küchler et al., 1999). Since then it has become a popular tool in stock price modeling and market risk measurement (Bibby and Sørensen, 2003; Chen, Härdle and Jeong, 2008; Eberlein, Keller and Prause, 1998; McNeil, Rüdiger and Embrechts, 2005).

The origin of the distribution dates back to the 1940s and the empirical observation by Ralph Bagnold that the log-histogram of the size of sand particles tends to form a hyperbola. A formal mathematical description was developed by Barndorff-Nielsen (1977). The hyperbolic distribution provides the possibility of modeling heavier tails than the Gaussian, since its log-density forms a hyperbola while that of the Gaussian is a parabola, see Figure 4. As we will see later in this Section, the hyperbolic law is a member of a larger, versatile class of generalized hyperbolic (GH) distributions, which also includes the normal-inverse Gaussian (NIG) and variance-gamma (VG) distributions as special cases. For a concise review of special and limiting cases of the GH distribution see Paoletta (2007), Chapter 9.

3.1.1 The hyperbolic distribution

The hyperbolic distribution is defined as a normal variance-mean mixture where the mixing distribution is the generalized inverse Gaussian (GIG) law with parameter $\lambda = 1$, i.e. it is conditionally Gaussian, see Barndorff-Nielsen (1977) and Barndorff-

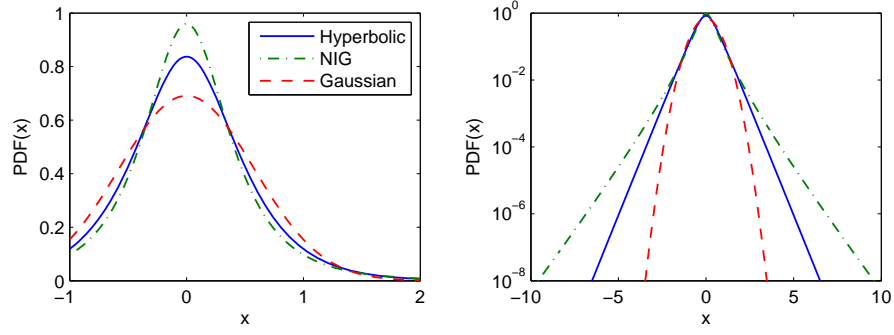


Fig. 4 Densities and log-densities of symmetric hyperbolic, NIG and Gaussian distributions having the same variance, see (31). The name of the hyperbolic distribution is derived from the fact that its log-density forms a hyperbola, which is clearly visible in the right panel.

Nielsen and Blaesild (1981). More precisely, a random variable Z has the hyperbolic distribution if:

$$(Z|Y) \sim N(\mu + \beta Y, Y), \quad (17)$$

where Y is a generalized inverse Gaussian $GIG(\lambda = 1, \chi, \psi)$ random variable and $N(m, s^2)$ denotes the Gaussian distribution with mean m and variance s^2 . The GIG law is a very versatile positive domain distribution with the PDF given by:

$$f_{GIG}(x) = \frac{(\psi/\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\chi\psi})} x^{\lambda-1} e^{-\frac{1}{2}(\chi x^{-1} + \psi x)}, \quad x > 0, \quad (18)$$

where the three parameters take values in one of the ranges: (i) $\chi > 0, \psi \geq 0$ if $\lambda < 0$, (ii) $\chi > 0, \psi > 0$ if $\lambda = 0$ or (iii) $\chi \geq 0, \psi = 0$ if $\lambda > 0$. The generalized inverse Gaussian law has a number of interesting properties that we will use later in this section. The distribution of the inverse of a GIG variable is again GIG but with a different λ , namely if:

$$Y \sim GIG(\lambda, \chi, \psi) \quad \text{then} \quad Y^{-1} \sim GIG(-\lambda, \chi, \psi). \quad (19)$$

A GIG variable can be also reparameterized by setting $a = \sqrt{\chi/\psi}$ and $b = \sqrt{\chi\psi}$, and defining $Y = a\tilde{Y}$, where:

$$\tilde{Y} \sim GIG(\lambda, b, b). \quad (20)$$

The normalizing constant $K_\lambda(t)$ in formula (18) is the modified Bessel function of the third kind with index λ , also known as the MacDonald function. It is defined as:

$$K_\lambda(t) = \frac{1}{2} \int_0^\infty x^{\lambda-1} e^{-\frac{1}{2}t(x+x^{-1})} dx, \quad t > 0. \quad (21)$$

In the context of hyperbolic distributions, the Bessel functions are thoroughly discussed in Barndorff-Nielsen and Blaesild (1981). Here we recall only two properties that will be used later. Namely, (i) $K_\lambda(t)$ is symmetric with respect to λ , i.e. $K_\lambda(t) = K_{-\lambda}(t)$, and (ii) for $\lambda = \pm\frac{1}{2}$ it can be written in a simpler form:

$$K_{\pm\frac{1}{2}}(t) = \sqrt{\frac{\pi}{2}} t^{-\frac{1}{2}} e^{-t}. \quad (22)$$

For other values of λ numerical approximations of the integral in eqn. (21) have to be used, see e.g. Press et al. (1992).

Relation (17) implies that a hyperbolic random variable $Z \sim H(\psi, \beta, \chi, \mu)$ can be represented in the form: $Z \sim \mu + \beta Y + \sqrt{Y}N(0, 1)$, with the CF:

$$\phi_Z(u) = e^{i\mu u} \int_0^\infty e^{i\beta zu - \frac{1}{2}zu^2} dF_Y(z). \quad (23)$$

Here $F_Y(z)$ denotes the distribution function of a GIG random variable Y with parameter $\lambda = 1$, see eqn. (18). Hence, the hyperbolic PDF is given by:

$$f_H(x; \psi, \beta, \chi, \mu) = \frac{\sqrt{\psi/\chi}}{2\sqrt{\psi + \beta^2} K_1(\sqrt{\psi\chi})} e^{-\sqrt{\{\psi + \beta^2\}\{\chi + (x-\mu)^2\}} + \beta(x-\mu)}, \quad (24)$$

or in an alternative parameterization (with $\delta = \sqrt{\chi}$ and $\alpha = \sqrt{\psi + \beta^2}$) by:

$$f_H(x; \alpha, \beta, \delta, \mu) = \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2 - \beta^2})} e^{-\alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu)}. \quad (25)$$

The latter is more popular and has the advantage of $\delta > 0$ being the traditional scale parameter. Out of the remaining three parameters, α and β determine the shape, with α being responsible for the steepness and $0 \leq |\beta| < \alpha$ for the skewness, and $\mu \in \mathbb{R}$ is the location parameter.

Finally, note that if we only have an efficient algorithm to compute K_1 , the calculation of the PDF is straightforward. However, the CDF has to be numerically integrated from (24) or (25).

3.1.2 The general class

The generalized hyperbolic (GH) law can be represented as a normal variance-mean mixture where the mixing distribution is the generalized inverse Gaussian law with any $\lambda \in \mathbb{R}$. Hence, the GH distribution is described by five parameters $\theta = (\lambda, \alpha, \beta, \delta, \mu)$, using parameterization (25), and its PDF is given by:

$$f_{GH}(x; \theta) = \kappa \{\delta^2 + (x - \mu)^2\}^{\frac{1}{2}(\lambda - \frac{1}{2})} K_{\lambda - \frac{1}{2}} \left(\alpha \sqrt{\delta^2 + (x - \mu)^2} \right) e^{\beta(x - \mu)}, \quad (26)$$

where:

$$\kappa = \frac{(\alpha^2 - \beta^2)^{\frac{\lambda}{2}}}{\sqrt{2\pi} \alpha^{\lambda - \frac{1}{2}} \delta^\lambda K_\lambda(\delta \sqrt{\alpha^2 - \beta^2})}. \quad (27)$$

The tail behavior of the GH density is ‘semi-heavy’, i.e. the tails are lighter than those of non-Gaussian stable laws, but much heavier than Gaussian. Formally they are characterized by the following asymptotic relation (Barndorff-Nielsen and Blaesild, 1981):

$$f_{\text{GH}}(x) \approx |x|^{\lambda-1} e^{(\mp\alpha+\beta)x} \quad \text{for } x \rightarrow \pm\infty, \quad (28)$$

which can be interpreted as exponential ‘tempering’ of the power-law tails (compare with the TSD described in Section 2.2).

For $|\beta + z| < \alpha$ the moment generating function of the GH law takes the form:

$$M(z) = e^{\mu z} \left\{ \frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + z)^2} \right\}^{\frac{\lambda}{2}} \frac{K_\lambda(\delta \sqrt{\alpha^2 - (\beta + z)^2})}{K_\lambda(\delta \sqrt{\alpha^2 - \beta^2})}. \quad (29)$$

Note, that $M(z)$ is smooth, i.e. infinitely many times differentiable, near 0 and hence every moment exists. If we set $\zeta = \delta \sqrt{\alpha^2 - \beta^2} = \sqrt{\psi\chi}$ then the first two moments lead to the following formulas for the mean and variance of a GH random variable:

$$\mathbb{E}X = \mu + \frac{\beta \delta^2}{\zeta} \frac{K_{\lambda+1}(\zeta)}{K_\lambda(\zeta)}, \quad (30)$$

$$\text{Var}X = \delta^2 \left[\frac{K_{\lambda+1}(\zeta)}{\zeta K_\lambda(\zeta)} + \frac{\beta^2 \delta^2}{\zeta^2} \left\{ \frac{K_{\lambda+2}(\zeta)}{K_\lambda(\zeta)} - \left(\frac{K_{\lambda+1}(\zeta)}{\zeta K_\lambda(\zeta)} \right)^2 \right\} \right]. \quad (31)$$

3.1.3 The normal-inverse Gaussian distribution

The normal-inverse Gaussian (NIG) laws were introduced by Barndorff-Nielsen (1995) as a subclass of the generalized hyperbolic laws obtained for $\lambda = -\frac{1}{2}$. The density of the NIG distribution is given by:

$$f_{\text{NIG}}(x) = \frac{\alpha \delta}{\pi} e^{\delta \sqrt{\alpha^2 - \beta^2} + \beta(x - \mu)} \frac{K_1(\alpha \sqrt{\delta^2 + (x - \mu)^2})}{\sqrt{\delta^2 + (x - \mu)^2}}. \quad (32)$$

Like for the hyperbolic distribution the calculation of the PDF is straightforward, but the CDF has to be numerically integrated from eqn. (32).

At the expense of four parameters, the NIG distribution is able to model asymmetric distributions with ‘semi-heavy’ tails. However, if we let $\alpha \rightarrow 0$ the NIG distribution converges to the Cauchy distribution (with location parameter μ and scale parameter δ), which exhibits extremely heavy tails. Obviously, the NIG distribution may not be adequate to deal with cases of extremely heavy tails such as those of Pareto or non-Gaussian stable laws. However, empirical experience

suggests excellent fits of the NIG law to financial data (Karlís, 2002; Karlís and Liljestöl, 2004; Venter and de Jongh, 2002).

Moreover, the class of normal-inverse Gaussian distributions possesses an appealing feature that the class of hyperbolic laws does not have. Namely, it is closed under convolution, i.e. a sum of two independent NIG random variables is again NIG (Barndorff-Nielsen, 1995). In particular, if X_1 and X_2 are independent NIG random variables with common parameters α and β but having different scale and location parameters $\delta_{1,2}$ and $\mu_{1,2}$, respectively, then $X = X_1 + X_2$ is $\text{NIG}(\alpha, \beta, \delta_1 + \delta_2, \mu_1 + \mu_2)$. This feature is especially useful in time scaling of risks, e.g. in deriving 10-day risks from daily risks. Only two subclasses of the generalized hyperbolic distributions are closed under convolution. The other class with this important property is the class of variance-gamma (VG) distributions, which is obtained when δ is equal to 0. This is only possible for $\lambda > 0$ and $\alpha > |\beta|$. The VG distributions (with $\beta = 0$) were introduced to finance by Madan and Seneta (1990), long before the popularity of GH and NIG laws.

3.2 Simulation of generalized hyperbolic variables

The most natural way of simulating GH variables stems from the fact that they can be represented as normal variance-mean mixtures. Since the mixing distribution is the GIG law, the resulting algorithm reads as follows:

1. simulate a random variable $Y \sim \text{GIG}(\lambda, \chi, \psi) = \text{GIG}(\lambda, \delta^2, \alpha^2 - \beta^2)$;
2. simulate a standard normal random variable N ;
3. return $X = \mu + \beta Y + \sqrt{Y}N$.

The algorithm is fast and efficient if we have a handy way of simulating GIG variates. For $\lambda = -\frac{1}{2}$, i.e. when sampling from the so-called inverse Gaussian (IG) distribution, there exists an efficient procedure that utilizes a transformation yielding two roots. It starts with the observation that if we let $\vartheta = \sqrt{\chi/\psi}$ then the $\text{IG}(\chi, \psi)$ density ($= \text{GIG}(-\frac{1}{2}, \chi, \psi)$; see eqn. (18)) of Y can be written as:

$$f_{\text{IG}}(x) = \sqrt{\frac{\chi}{2\pi x^3}} \exp \left\{ \frac{-\chi(x - \vartheta)^2}{2x\vartheta^2} \right\}. \quad (33)$$

Now, following Shuster (1968) we may write:

$$V = \frac{\chi(Y - \vartheta)^2}{Y\vartheta^2} \sim \chi_{(1)}^2, \quad (34)$$

i.e. V is distributed as a chi-square random variable with one degree of freedom. As such it can be simply generated by taking a square of a standard normal random number. Unfortunately, the value of Y is not uniquely determined by eqn. (34). Solving this equation for Y yields two roots:

$$y_1 = \vartheta + \frac{\vartheta}{2\chi} \left(\vartheta V - \sqrt{4\vartheta\chi V + \vartheta^2 V^2} \right) \quad \text{and} \quad y_2 = \frac{\vartheta^2}{y_1}.$$

The difficulty in generating observations with the desired distribution now lies in choosing between the two roots. Michael, Schucany and Haas (1976) showed that Y can be simulated by choosing y_1 with probability $\vartheta/(\vartheta + y_1)$. So for each random observation V from a $\chi_{(1)}^2$ distribution the smaller root y_1 has to be calculated. Then an auxiliary Bernoulli trial is performed with probability $p = \vartheta/(\vartheta + y_1)$. If the trial results in a ‘success’, y_1 is chosen; otherwise, the larger root y_2 is selected. The *rnig* function of the Rmetrics collection of software packages for S-plus/R (see also Section 2.3 where Rmetrics was briefly described), utilize this routine.

In the general case, the GIG distribution – as well as the (generalized) hyperbolic law – can be simulated via the rejection algorithm. An adaptive version of this technique is used to obtain hyperbolic random numbers in the *rhyp* function of Rmetrics. Rejection is also implemented in the HyperbolicDist package for S-plus/R developed by David Scott, see the R-project home page cran.r-project.org. The package utilizes a version of the algorithm proposed by Atkinson (1982), i.e. rejection coupled either with a two (‘GIG algorithm’ for any admissible value of λ) or a three part envelope (‘GIGLT1 algorithm’ for $0 \leq \lambda < 1$). Envelopes, also called hat or majorizing functions, provide an upper limit for the PDF of the sampled distribution. The proper choice of such functions can substantially increase the speed of computations, see Chapter ?? . As Atkinson (1982) shows, once the parameter values for these envelopes have been determined, the algorithm efficiency is reasonable for most values of the parameter space. However, finding the appropriate parameters requires optimization and makes the technique burdensome.

This difficulty led to a search for a short algorithm which would give comparable efficiencies but without the drawback of extensive numerical optimizations. A solution, based on the ‘ratio-of-uniforms’ method, was provided by Dagpunar (1989). First, recalling properties (19) and (20), observe that we only need to find a method to simulate $\tilde{Y} \sim \text{GIG}(\lambda, b, b)$ variables and only for $\lambda \geq 0$. Next, define the relocated variable $\tilde{Y}_m = \tilde{Y} - m$, where $m = \frac{1}{b}(\lambda - 1 + \sqrt{(\lambda - 1)^2 + b^2})$ is the mode of the density of \tilde{Y} . Then the relocated variable can be generated by taking $\tilde{Y}_m = V/U$, where the pair (U, V) is uniformly distributed over the region $\{(u, v) : 0 \leq u \leq \sqrt{h(v/u)}\}$ with:

$$h(t) = (t + m)^{\lambda-1} \exp\left(-\frac{b}{2} \frac{t + m + 1}{t + m}\right), \quad \text{for } t \geq -m.$$

Since this region is irregularly shaped, it is more convenient to generate the pair (U, V) uniformly over a minimal enclosing rectangle $\{(u, v) : 0 \leq u \leq u_+, v_- \leq v \leq v_+\}$. Finally, the variate (V/U) is accepted if $U^2 \leq h(V/U)$. The efficiency of the algorithm depends on the method of deriving and the actual choice of u_+ and v_{\pm} . Further, for $\lambda \leq 1$ and $b \leq 1$ there is no need for the shift at mode m . Such a version of the algorithm is implemented in UNU.RAN, a library of C functions for non-uniform random number generation developed at the Vienna University of Economics, see statistik.wu-wien.ac.at/unuran.

3.3 Estimation of parameters

3.3.1 Maximum likelihood method

The parameter estimation of GH distributions can be performed by the ML method, since there exist closed-form formulas (although, involving special functions) for the densities of these laws. The computational burden is not as heavy as for stable laws, but it still is considerable.

In general, the ML estimation algorithm is as follows. For a vector of observations $x = (x_1, \dots, x_n)$, the ML estimate of the parameter vector $\theta = (\lambda, \alpha, \beta, \delta, \mu)$ is obtained by maximizing the log-likelihood function:

$$L(x; \theta) = \log \kappa + \frac{\lambda - \frac{1}{2}}{2} \sum_{i=1}^n \log(\delta^2 + (x_i - \mu)^2) + \sum_{i=1}^n \log K_{\lambda - \frac{1}{2}}(\alpha \sqrt{\delta^2 + (x_i - \mu)^2}) + \sum_{i=1}^n \beta(x_i - \mu), \quad (35)$$

where κ is defined by (27). Obviously, for hyperbolic ($\lambda = 1$) distributions the algorithm uses simpler expressions of the log-likelihood function due to relation (22).

The routines proposed in the literature differ in the choice of the optimization scheme. The first software product that allowed statistical inference with hyperbolic distributions – the HYP program – used a gradient search technique, see Blaesild and Sorensen (1992). In a large simulation study Prause (1999) utilized the bracketing method. Matlab functions *hypst.m* and *nigest.m* distributed with the MFE Toolbox (Weron, 2006) use yet another technique – the downhill simplex method, with slight modifications due to parameter restrictions.

The main factor for the speed of the estimation is the number of modified Bessel functions to compute. Note, that for $\lambda = 1$ (i.e. the hyperbolic distribution) this function appears only in the constant κ . For a data set with n independent observations we need to evaluate n and $n + 1$ Bessel functions for NIG and generalized hyperbolic distributions, respectively, whereas only one for the hyperbolic. This leads to a considerable reduction in the time necessary to calculate the likelihood function in the hyperbolic case. Prause (1999) reported a reduction of ca. 33%, however, the efficiency results are highly sample and implementation dependent.

We also have to say that the optimization is challenging. Some of the parameters are hard to separate since a flat-tailed GH distribution with a large scale parameter is hard to distinguish from a fat-tailed distribution with a small scale parameter, see Barndorff-Nielsen and Blaesild (1981) who observed such a behavior already for the hyperbolic law. The likelihood function with respect to these parameters then becomes very flat, and may have local minima. In the case of NIG distributions Venter and de Jongh (2002) proposed simple estimates of α and β that can be used as starting values for the ML scheme. Starting from relation (28) for the tails of the NIG density (i.e. with $\lambda = -1/2$) they derived the following approximation:

$$\alpha - \beta \sim \frac{1}{2} \frac{x_{1-f} + \mathbb{E}(X|X > x_{1-f})}{\mathbb{E}(X^2|X > x_{1-f}) - x_{1-f}\mathbb{E}(X|X > x_{1-f})},$$

$$\alpha + \beta \sim -\frac{1}{2} \frac{x_f + \mathbb{E}(X|X < x_f)}{\mathbb{E}(X^2|X < x_f) - x_f\mathbb{E}(X|X < x_f)},$$

where x_f is the f -th population quantile, see Section 2.5. After the choice of a suitable value for f , Venter and de Jongh used $f = 5\%$, the ‘tail estimates’ of α and β are obtained by replacing the quantiles and expectations by their sample values in the above relations.

Another method of providing the starting values for the ML scheme was suggested by Prause (1999). He estimated the parameters of a symmetric ($\beta = \mu = 0$) GH law with a reasonable kurtosis (i.e. with $\delta\alpha \approx 1.04$) that had the variance equal to that of the empirical distribution.

3.3.2 Other methods

Besides the ML approach other estimation methods have been proposed in the literature. Prause (1999) tested different estimation techniques by replacing the log-likelihood function with other score functions, like the Anderson-Darling and Kolmogorov statistics or L^p -norms. But the results were disappointing. Karlis and Lillstöl (2004) made use of the MCMC technique (see Chapter ??), however, again the results obtained were not impressive. Karlis (2002) described an Expectation-Maximization (EM) type algorithm (see Chapter ??) for ML estimation of the NIG distribution. The algorithm can be programmed in any statistical package supporting Bessel functions and it has all the properties of the standard EM algorithm, like sure, but slow, convergence, parameters in the admissible range, etc. Recently Fragiadakis, Karlis and Meintanis (2009) used this approach to construct goodness-of-fit tests for symmetric NIG distributions. The tests are based on a weighted integral incorporating the empirical CF of suitably standardized data. The EM scheme can be also generalized to multivariate GH distributions (but with fixed λ , see Protassov, 2004).

4 Value at Risk, portfolios and heavy tails

4.1 Copulas

The facts presented in Section 1 clearly show that we not only can, but must use heavy tailed alternatives to the Gaussian law in order to obtain acceptable estimates of market losses. But can we substitute the Gaussian distribution with other distributions in Value at Risk (Expected Shortfall) calculations for whole portfolios of assets? Recall, that the definition of VaR utilizes the quantiles of the portfolio returns

distribution and not the returns distribution of individual assets in the portfolio. If all asset return distributions are assumed to be Gaussian and linearly dependent then the portfolio distribution is multivariate normal and well known statistical tools can be applied (Härdle and Simar, 2003). However, when asset returns are distributed according to a different law (or different laws!) then the multivariate distribution may be hard to tackle. In particular, linear correlation may no longer be a meaningful measure of dependence.

It turns out that in this context the concept of copulas is very helpful (Joe, 1997; Nelsen, 1999). In rough terms, a copula is a multivariate distribution function defined on the unit cube $[0, 1]^n$, i.e. $C : [0, 1]^n \rightarrow [0, 1]$. What is important for VaR calculations is that a copula enables us to construct a multivariate distribution function from the marginal (possibly different) distribution functions of n individual asset returns in a way that takes their dependence structure into account. This dependence structure may be no longer measured by correlation, but by other adequate functions like rank correlation, comonotonicity or tail dependence (McNeil, Rüdiger and Embrechts, 2005). Moreover, it can be shown that for every multivariate distribution function there exists a copula which contains all information on dependence (this is the essence of ‘Sklar’s theorem’). For example, if the random variables are independent, then the independence (or product) copula is just the product of n variables: $C(u_1, \dots, u_n) = u_1 \cdot \dots \cdot u_n$. If the random variables have a multivariate normal distribution with a given covariance matrix then the Gaussian copula is obtained.

Copula functions do not impose any restrictions on the model, so in order to reach a model that is to be useful in a given risk management problem, a particular specification of the copula must be chosen. From the wide variety of copulas that exist probably the elliptical and Archimedean copulas are the ones most often used in applications. Elliptical copulas are simply the copulas of elliptically contoured (or elliptical) distributions, e.g. (multivariate) normal, t , symmetric stable and symmetric generalized hyperbolic (Fang, Kotz and Ng, 1987). Rank correlation and tail dependence coefficients can be easily calculated for elliptical copulas. There are, however, drawbacks – elliptical copulas do not have closed form expressions, are restricted to have radial symmetry and have all marginal distributions of the same type. These restrictions may disqualify elliptical copulas from being used in some risk management problems. In particular, there is usually a stronger dependence between big losses (e.g. market crashes) than between big gains. Clearly, such asymmetries cannot be modeled with elliptical copulas. In contrast to elliptical copulas, all commonly encountered Archimedean copulas have closed form expressions. Their popularity also stems from the fact that they allow for a great variety of different dependence structures. Many interesting parametric families of copulas are Archimedean, including the well known Clayton:

$$C_C(u_1, u_2; \theta) = \left(u_1^{-\theta} + u_2^{-\theta} - 1 \right)^{-1/\theta}, \quad (36)$$

and Frank:

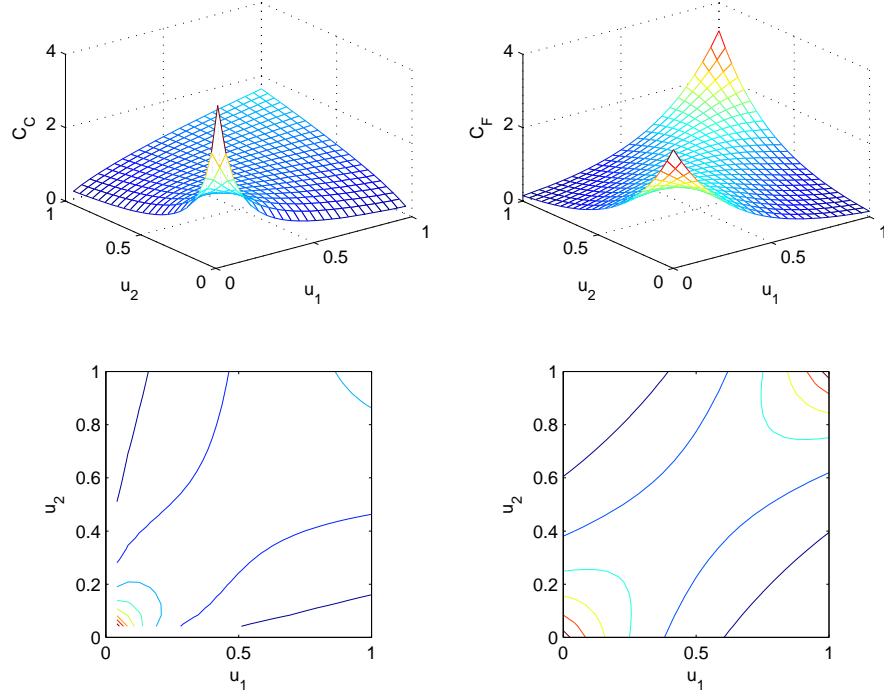


Fig. 5 Densities of a two-dimensional Clayton copula C_C with $\theta_C = 0.6515$ (*top left*) and a two-dimensional Frank copula C_F with $\theta_F = 3.0884$ (*top right*). *Bottom panels*: Contour plots of the densities in the upper panels. Note, that with this choice of the dependence parameters (θ_C, θ_F) the Clayton copula models strong dependence between big losses ($u_1, u_2 \rightarrow 0$), while the Frank copula focuses on both, big losses and big gains.

$$C_F(u_1, u_2; \theta) = -\frac{1}{\theta} \log \left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right), \quad (37)$$

copulas. In Figure 5 their densities are plotted for sample values of the dependence parameter θ corresponding to the estimates obtained for the portfolio considered in Section 4.2.

4.1.1 Calibration

Calibration of multivariate copula-based models can be performed in a number of ways. Simultaneous estimation of all parameters (of the copula and marginal distributions) using the full (or exact) maximum likelihood (FML, EML) approach is the most direct estimation method. However, this comes at the price of solving a complicated, multivariate and typically nonlinear optimization problem.

A feasible alternative which exploits the attractive feature of copulas for which the dependence structure is independent of the marginal distributions is known as the Inference Functions for Margins (IFM, see Fusai and Roncoroni, 2008; Joe, 1997) or the sequential two-step maximum likelihood (TSML, see Trivedi and Zimmer, 2005) method. In this approach, first the marginals are estimated on univariate data, then the dependence parameter θ is calibrated after the estimated marginals have been substituted into the copula. The IFM method has additional variants depending upon whether the first step is implemented parametrically or nonparametrically (using kernel density estimates). For a bivariate copula the procedure could be summarized as follows:

1. Using ML estimate the univariate marginal densities $\hat{f}_j(\mathbf{x}_j)$, $j = 1, 2$, for two random i.i.d. samples $\mathbf{x}_j = \{x_{j,1}, \dots, x_{j,N}\}$. In the nonparametric variant a kernel density estimator is used to find the \hat{f}_j 's.
2. Set $\mathbf{u}_j = \hat{F}_j(\mathbf{x}_j)$, where F_j is the CDF corresponding to f_j . The vectors \mathbf{u}_j may be treated as realizations of uniform random variables.
3. Given \mathbf{u}_j , $j = 1, 2$, and a copula C with density c , the dependence parameter θ can be estimated as follows:

$$\hat{\theta} = \arg \max_{\theta} \sum_{i=1}^N \log c(\hat{u}_{1,i}, \hat{u}_{2,i}; \theta). \quad (38)$$

4.1.2 Simulating random variables from copulas

For risk management purposes, we are interested in the Value at Risk of a portfolio of assets. While analytical methods for the computation of VaR exist for the multivariate normal distribution (i.e. for the Gaussian copula), in most other cases we have to use Monte Carlo simulations. A general technique for random variate generation from copulas is the conditional distributions method (Nelsen, 1999), also called conditional sampling (Fusai and Roncoroni, 2008). A random vector $(u_1, \dots, u_n)^T$ having a joint distribution function C can be generated by the following algorithm:

1. simulate $u_1 \sim U(0, 1)$,
2. for $k = 2, \dots, n$ simulate $u_k \sim F_{U_k|U_1 \dots U_{k-1}}(\cdot | u_1, \dots, u_{k-1})$.

The above function is the conditional distribution of the variable U_k given the values of U_1, \dots, U_{k-1} , i.e.:

$$\begin{aligned} F_{U_k|U_1 \dots U_{k-1}}(u_k | u_1, \dots, u_{k-1}) &\stackrel{\text{def}}{=} \mathbb{P}(U_k \leq u_k | U_1 = u_1, \dots, U_{k-1} = u_{k-1}) \\ &= \frac{\partial^{k-1} C_k(u_1, \dots, u_k)}{\partial u_1 \dots \partial u_{k-1}} \bigg/ \frac{\partial^{k-1} C_{k-1}(u_1, \dots, u_{k-1})}{\partial u_1 \dots \partial u_{k-1}}, \end{aligned}$$

where C_k 's are k -dimensional margins of the n -dimensional copula C , i.e.

$$C_k(u_1, \dots, u_k) = C(u_1, \dots, u_k, 1, \dots, 1).$$

The main drawback of this method is the fact that it involves a differentiation step for each dimension of the problem. Also simulation of $u_k \sim F_{U_k|U_1 \dots U_{k-1}}(\cdot|u_1, \dots, u_{k-1})$ may be non-trivial. It requires drawing v from a uniform distribution and setting $u_k \sim F_{U_k|U_1 \dots U_{k-1}}^{-1}(v|u_1, \dots, u_{k-1})$, with often having to compute F^{-1} numerically. Hence, the conditional distributions technique is typically not practical in higher dimensions. For this reason, alternative methods have been developed for specific types of copulas. For instance, random variables distributed according to Archimedean copula functions can be generated by the mixture of powers method of Marshall and Olkin (1988), which utilizes Laplace transforms. Also conditional sampling can be simplified for Archimedean copulas with the conditional distribution $F_{U_k|U_1 \dots U_{k-1}}$ rewritten in terms of the copula generator function. A comprehensive list of algorithms can be found in Alexander (2008), Fusai and Roncoroni (2008), and McNeil, Rüdiger and Embrechts (2005).

4.2 Empirical evidence

In this section we apply the techniques discussed so far to financial data. We want to build a VaR model for a hypothetical portfolio consisting (in equal parts) of the Deutsche Aktienindex (DAX) and the Polish WIG20 index. Both are blue chip stock market indexes. DAX consists of the 30 major German companies trading on the Frankfurt Stock Exchange and WIG20 of the 20 major Polish companies trading on the Warsaw Stock Exchange. We use daily closing index values from the period January 3, 2000 – December 31, 2009. Eliminating missing values (mostly German and Polish holidays) we end up with 2494 (log-)returns for each index, see the top left panels in Figures 6 and 7.

4.2.1 Building a VaR model

Like most financial time series, also these index returns contain volatility clusters which imply that the probability of a specific incurred loss is not the same on each day. During days of higher volatility we should expect larger than usual losses and during calmer days – smaller than usual. To remove volatility clusters it is necessary to model the process that generates them. Following Barone-Adesi, Giannopoulos, and Vosper (1999) and Kuester, Mittnik, and Paoletta (2006) we eliminate volatility clusters by filtering the returns r_t with a GARCH(1,1) process:

$$r_t = \sigma_t \varepsilon_t, \quad \text{with} \quad \sigma_t = c_0 + c_1 r_{t-1}^2 + d_1 \sigma_{t-1}^2, \quad (39)$$

where c_0 , c_1 and d_1 are constants and

$$\varepsilon_t = \frac{r_t}{\sigma_t}, \quad (40)$$

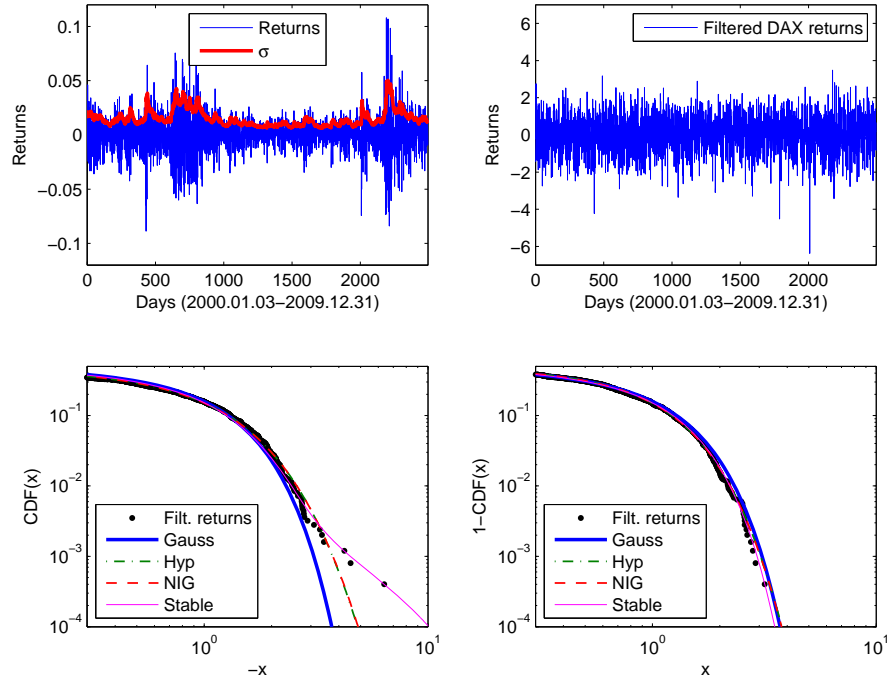


Fig. 6 *Top left:* DAX index (log-)returns and the GARCH(1,1)-based daily volatility estimate σ_t . *Top right:* σ_t -filtered DAX returns, see eqn. (40). *Bottom panels:* The left and right tails of the CDF of filtered returns and of four fitted distributions: Gaussian, hyperbolic, NIG, and stable (for parameter estimates see Table 2). Note, that the left tail is significantly heavier than the right tail. The latter is pretty well modeled by the Gaussian law.

Table 2 Gaussian, hyperbolic, NIG and stable fits to 2494 filtered returns of the Deutsche Aktienindex (DAX) from the period January 3, 2000 – December 31, 2009. The values of the Kolmogorov (K) and Anderson-Darling (AD) goodness-of-fit statistics suggest the hyperbolic distribution as the best model, with the NIG law following closely by.

Distribution	Parameters				Statistics	
	α	σ (δ)	β	μ	K	AD
Gaussian		1.000		-0.007	1.858	3.554
Hyperbolic	2.769	2.009	-0.542	0.508	0.779	0.623
NIG	2.536	2.350	-0.555	0.520	0.796	0.643
Stable	1.957	0.682	-1.000	-0.019	1.389	1.697

are the filtered returns, see the top right panels in Figures 6 and 7. We could also insert a moving average term in the conditional mean to remove any serial dependency if needed.

To find the right model class for each dataset we fit four distributions: Gaussian, hyperbolic, NIG, and stable to the DAX and WIG20 filtered returns. The results

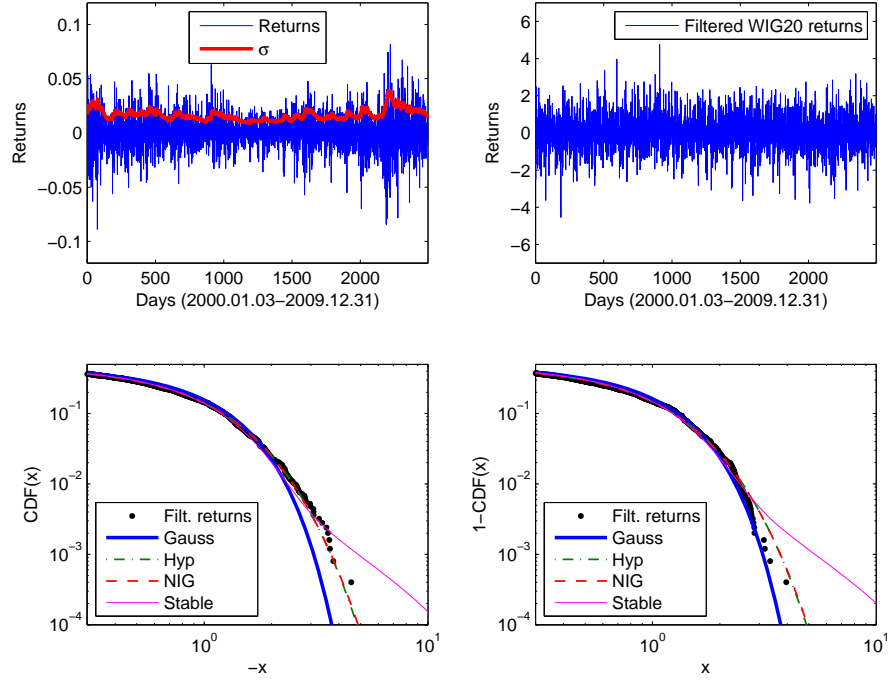


Fig. 7 *Top left*: WIG20 index (log-)returns and the GARCH(1,1)-based daily volatility estimate σ_t . *Top right*: σ_t -filtered WIG20 returns, see eqn. (40). *Bottom panels*: The left and right tails of the CDF of filtered returns and of four fitted distributions: Gaussian, hyperbolic, NIG, and stable (for parameter estimates see Table 3). Unlike for the DAX returns, the left tail is not significantly heavier than the right tail.

Table 3 Gaussian, hyperbolic, NIG and stable fits to 2494 filtered returns of the Polish WIG20 index from the period January 3, 2000 – December 31, 2009. Like for the filtered DAX returns, the values of the Kolmogorov (K) and Anderson-Darling (AD) goodness-of-fit statistics suggest the hyperbolic distribution as the best model, with the NIG law following closely by.

Distribution	Parameters				Statistics	
	α	σ (δ)	β	μ	K	AD
Gaussian		1.003		0.008	1.584	3.500
Hyperbolic	2.048	1.284	-0.002	0.010	0.681	0.452
NIG	1.732	1.747	0.030	-0.022	0.682	0.463
Stable	1.890	0.654	0.207	0.020	0.881	1.095

are presented in Tables 2 and 3. We also compare both fits using Kolmogorov (K) and Anderson-Darling (AD) test statistics (D’Agostino and Stephens, 1986). The latter may be treated as a weighted Kolmogorov statistics which puts more weight to the differences in the tails of the distributions. Although no asymptotic results are known for stable or generalized hyperbolic laws, approximate critical values for these goodness-of-fit tests can be obtained via the bootstrap technique (Cizek, Härdle and Weron, 2005; Stute, Manteiga and Quindimil, 1993). In this chapter, though, we will not perform hypothesis testing and just compare the test values. Naturally, the lower the values the better the fit. For both datasets, both statistics suggest the hyperbolic distribution as the best model, with the NIG law following closely by. Note, that for the DAX filtered returns the left tail is significantly heavier than the right tail, with the latter being pretty well modeled by the Gaussian law, see the bottom panels in Figures 6 and 7. This is also confirmed by very negative skewness parameters (β). In contrast, the WIG20 filtered returns are roughly symmetric.

4.2.2 Computing VaR and backtesting the models

Based on the goodness-of-fit results of Section 4.2.1 we use two distributional classes for the marginals – hyperbolic or NIG (for simplicity the same distributional class is used for both marginals). The dependence structure is modeled either by the Clayton or the Frank copula. This leaves us with four models: Clayton(Hyp,Hyp), Clayton(NIG,NIG), Frank(Hyp,Hyp), and Frank(NIG,NIG). As a benchmark model we use the Filtered Historical Simulation (FHS) introduced by Barone-Adesi, Giannopoulos, and Vosper (1999), which has been documented to perform pretty well under different circumstances (Kuester, Mittnik, and Paolella, 2006).

The dynamic simulation scenario consists of computing the one day VaR at four different levels (90%, 95%, 97.5%, and 99%) for a rolling window of 500 observations (roughly two years of data). This leaves us with $2494 - 500 = 1994$ daily VaR estimates for each of the five methods, see Figure 8. These forecasts are then compared with the actual returns of the portfolio in the backtesting procedure. Each day is marked as 0 if VaR (at a given level) is not exceeded and 1 otherwise. The resulting sequence is Bernoulli distributed with parameter $(1 - c)$, see eqn. (1).

Several statistical tests have been proposed for backtesting purposes. In this study, we use Christoffersen’s (1998) approach to test the conditional coverage. This model independent approach is designed to overcome the clustering effect. The tests are carried out in the likelihood ratio (LR) framework. Three LR statistics are calculated: for the unconditional coverage, independence and conditional coverage. The former two are distributed asymptotically as $\chi^2(1)$ and the latter as $\chi^2(2)$. If we condition on the first observation, then the conditional coverage LR test statistics is the sum of the other two.

The p -values of Christoffersen’s (1998) conditional coverage test for the five considered VaR models are displayed in Table 4. The models based on the Frank copula produce disappointing results and fail at the high VaR levels. The Clayton(NIG,NIG) model yields the best coverage over all VaR levels. The Clay-

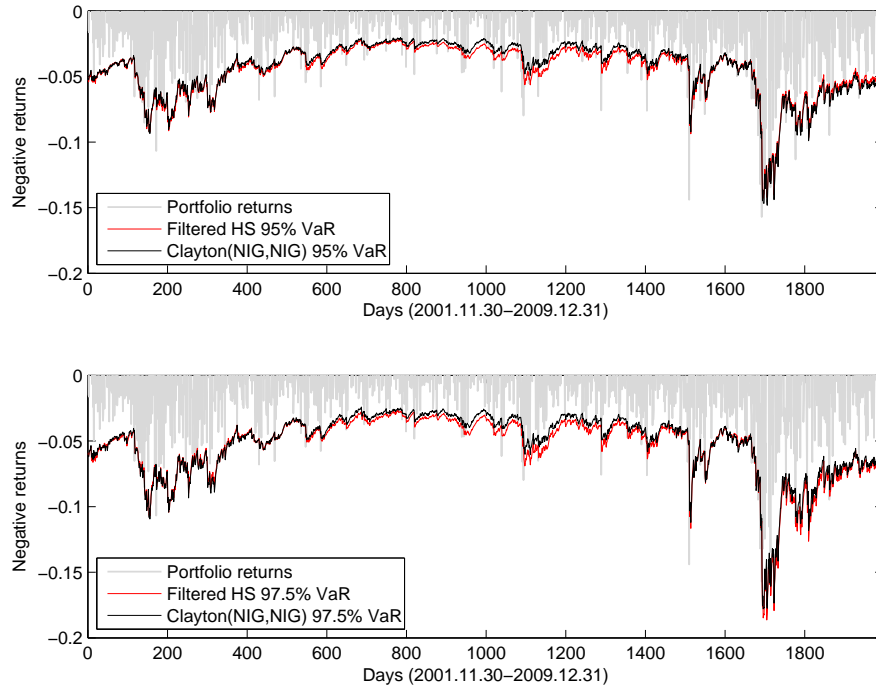


Fig. 8 The DAX-WIG20 portfolio returns and one day ahead $\text{VaR}_{95\%}$ (*top*) and $\text{VaR}_{97.5\%}$ (*bottom*) estimates for the FHS and Clayton(NIG,NIG) models. The latter yields slightly lower (in absolute terms) VaR numbers and a better conditional coverage, especially at the 97.5% level (see Table 4).

Table 4 The p -values of Christoffersen's (1998) conditional coverage test for the five considered VaR models at four different levels (90%, 95%, 97.5%, and 99%). The Clayton(NIG,NIG) model yields the best coverage over all VaR levels, with the Clayton(Hyp,Hyp) model and FHS following closely by. The models based on the Frank copula fail at the high VaR levels.

Model	$\text{VaR}_{90\%}$	$\text{VaR}_{95\%}$	$\text{VaR}_{97.5\%}$	$\text{VaR}_{99\%}$
FHS	0.360	0.719	0.441	0.778
Clayton(Hyp, Hyp)	0.462	0.715	0.953	0.685
Frank(Hyp, Hyp)	0.401	0.103	0.028	0.003
Clayton(NIG,NIG)	0.505	0.719	0.971	0.814
Frank(NIG,NIG)	0.401	0.148	0.021	0.001

ton(Hyp,Hyp) model is a little worse, despite a slightly better fit of the hyperbolic distribution to the filtered returns of the portfolio components, see Tables 2 and 3. The FHS model follows closely by and underperforms only at the 97.5% level. Perhaps, as Pritsker (2006) argues, two years of daily data (roughly 500 daily returns) may not contain enough extreme outliers to accurately compute 1% VaR.

Copulas allow us to construct models which go beyond the standard notions of correlation and multivariate Gaussian distributions. As such, in conjunction with alternative asset returns distributions discussed earlier in this chapter, they yield an ideal tool to model a wide variety of financial portfolios and products. No wonder they are gradually becoming an element of good risk management practice.

References

- Alexander, C. (2008). *Market Risk Analysis*, Wiley, Chichester.
- Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk, *Mathematical Finance* **9**: 203–228.
- Atkinson, A. C. (1982). The simulation of generalized inverse Gaussian and hyperbolic random variables, *SIAM Journal of Scientific & Statistical Computing* **3**: 502–515.
- Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size, *Proceedings of the Royal Society London A* **353**: 401–419.
- Barndorff-Nielsen, O. E. (1995). *Normal\Inverse Gaussian Processes and the Modelling of Stock Returns*, Research Report 300, Department of Theoretical Statistics, University of Aarhus.
- Barndorff-Nielsen, O. E. and Blaesild, P. (1981). Hyperbolic distributions and ramifications: Contributions to theory and applications, in C. Taillie, G. Patil, B. Baldessari (eds.) *Statistical Distributions in Scientific Work*, Volume 4, Reidel, Dordrecht, pp. 19–44.
- Barone-Adesi, G., Giannopoulos, K., and Vosper, L., (1999). VaR without correlations for portfolios of derivative securities, *Journal of Futures Markets* **19**(5): 583–602.
- Basle Committee on Banking Supervision (1995). An internal model-based approach to market risk capital requirements, <http://www.bis.org>.
- Bianchi, M. L., Rachev, S. T., Kim, Y. S. and Fabozzi, F. J. (2010). Tempered stable distributions and processes in finance: Numerical analysis, in M. Corazza, P. Claudio, P. (Eds.) *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, Springer.
- Bibby, B. M. and Sørensen, M. (2003). Hyperbolic processes in finance, in S. T. Rachev (ed.) *Handbook of Heavy-tailed Distributions in Finance*, North Holland.
- Blaesild, P. and Sørensen, M. (1992). *HYP – a Computer Program for Analyzing Data by Means of the Hyperbolic Distribution*, Research Report 248, Department of Theoretical Statistics, Aarhus University.
- Boyarchenko, S. I. and Levendorskii, S. Z. (2000). Option pricing for truncated Lévy processes, *International Journal of Theoretical and Applied Finance* **3**: 549–552.
- Breich, R. F., Iskander, D. R. and Zoubir, A. M. (2005). The stability test for symmetric alpha stable distributions, *IEEE Transactions on Signal Processing* **53**: 977–986.
- Buckle, D. J. (1995). Bayesian inference for stable distributions, *Journal of the American Statistical Association* **90**: 605–613.
- Carr, P., Geman, H., Madan, D. B. and Yor, M. (2002). The fine structure of asset returns: an empirical investigation, *Journal of Business* **75**: 305–332.
- Chambers, J. M., Mallows, C. L. and Stuck, B. W. (1976). A Method for Simulating Stable Random Variables, *Journal of the American Statistical Association* **71**: 340–344.
- Chen, Y., Härdle, W. and Jeong, S.-O. (2008). Nonparametric risk management with generalized hyperbolic distributions, *Journal of the American Statistical Association* **103**: 910–923.

- Christoffersen, P. (1998). Evaluating interval forecasts, *International Economic Review* **39**(4): 841–862.
- Cizek, P., Härdle, W. and Weron, R. (2005). *Statistical Tools in Finance and Insurance*, Springer.
- Cont, R., Potters, M. and Bouchaud, J.-P. (1997). Scaling in stock market data: Stable laws and beyond, in B. Dubrulle, F. Graner, D. Sornette (eds.) *Scale Invariance and Beyond*, Proceedings of the CNRS Workshop on Scale Invariance, Springer, Berlin.
- D’Agostino, R. B. and Stephens, M. A. (1986). *Goodness-of-Fit Techniques*, Marcel Dekker, New York.
- Dagpunar, J. S. (1989). An Easily Implemented Generalized Inverse Gaussian Generator, *Communications in Statistics – Simulations* **18**: 703–710.
- Danielsson, J., Hartmann, P. and De Vries, C. G. (1998). The cost of conservatism: Extreme returns, value at risk and the Basle multiplication factor, *Risk* **11**: 101–103.
- Dominicy, Y. and Veredas, D. (2010). The method of simulated quantiles, *ECARES working paper*, 2010–008.
- DuMouchel, W. H. (1971). *Stable Distributions in Statistical Inference*, Ph.D. Thesis, Department of Statistics, Yale University.
- DuMouchel, W. H. (1973). On the Asymptotic Normality of the Maximum–Likelihood Estimate when Sampling from a Stable Distribution, *Annals of Statistics* **1**(5): 948–957.
- Eberlein, E. and Keller, U. (1995). Hyperbolic distributions in finance, *Bernoulli* **1**: 281–299.
- Eberlein, E., Keller, U. and Prause, K. (1998). New insights into the smile, mispricing and Value at Risk: The hyperbolic model, *Journal of Business* **71**: 371–406.
- Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer.
- Fama, E. F. and Roll, R. (1971). Parameter Estimates for Symmetric Stable Distributions, *Journal of the American Statistical Association* **66**: 331–338.
- Fan, Z. (2006). Parameter estimation of stable distributions, *Communications in Statistics – Theory and Methods* **35**(2): 245–255.
- Fang, K.-T., Kotz, S. and Ng, K.-W. (1987). *Symmetric Multivariate and Related Distributions*, Chapman & Hall, London.
- Fofack, H. and Nolan, J. P. (1999). Tail Behavior, Modes and Other Characteristics of Stable Distributions, *Extremes* **2**: 39–58.
- Fragiadakis, K., Karlis, D. and Meintanis, S. G. (2009). Tests of fit for normal inverse Gaussian distributions, *Statistical Methodology* **6**: 553–564.
- Fusai, G., and Roncoroni, A. (2008). *Implementing Models in Quantitative Finance: Methods and Cases*, Springer.
- Garcia, R., Renault, E. and Veredas, D. (2010). Estimation of stable distributions by indirect inference, *Journal of Econometrics*, Forthcoming.
- Grabchak, M. (2010). Maximum likelihood estimation of parametric tempered stable distributions on the real line with applications to finance, *Ph.D. thesis*, Cornell University.
- Grabchak, M. and Samorodnitsky, G. (2010). Do financial returns have finite or infinite variance? A paradox and an explanation, *Quantitative Finance*, DOI: 10.1080/14697680903540381.
- Guillaume, D. M., Dacorogna, M. M., Dave, R. R., Müller, U. A., Olsen, R. B. and Pictet, O. V. (1997). From the birds eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets, *Finance & Stochastics* **1**: 95–129.
- Härdle, W., Klinke, S. and Müller, M. (2000). *XploRe Learning Guide*, Springer. See also: <http://www.xplo-re-stat.de/ebooks/ebooks.html>.
- Härdle, W., Kleinow, T. and Stahl, G. (2002). *Applied Quantitative Finance*, Springer. See also: <http://www.xplo-re-stat.de/ebooks/ebooks.html>.
- Härdle, W. and Simar, L. (2003). *Applied Multivariate Statistical Analysis*, Springer. See also: <http://www.xplo-re-stat.de/ebooks/ebooks.html>.
- Holt, D. R. and Crow, E. L. (1973). Tables and graphs of the stable probability density functions, *Journal of Research of the National Bureau of Standards B* **77B**: 143–198.
- Janicki, A. and Kokoszka, P. (1992). Computer investigation of the rate of convergence of LePage type series to alpha-stable random variables, *Statistica* **23**: 365–373.

- Janicki, A. and Weron, A. (1994a). Can one see α -stable variables and processes, *Statistical Science* **9**: 109–126.
- Janicki, A. and Weron, A. (1994b). *Simulation and Chaotic Behavior of α -Stable Stochastic Processes*, Marcel Dekker.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.
- Jorion, P. (2006). *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd ed., McGraw-Hill.
- Karlis, D. (2002). An EM type algorithm for maximum likelihood estimation for the Normal Inverse Gaussian distribution, *Statistics and Probability Letters* **57**: 43–52.
- Karlis, D. and Lillstöl, J. (2004). Bayesian estimation of NIG models via Markov chain Monte Carlo methods, *Applied Stochastic Models in Business and Industry* **20**(4): 323–338.
- Kawai, R. and Masuda, H. (2010). On simulation of tempered stable random variates, *Preprint*, Kyushu University.
- Khindarova, I., Rachev, S. and Schwartz, E. (2001). Stable Modeling of Value at Risk, *Mathematical and Computer Modelling* **34**: 1223–1259.
- Kogon, S. M. and Williams, D. B. (1998). Characteristic function based estimation of stable parameters, in R. Adler, R. Feldman, M. Taqqu (eds.), *A Practical Guide to Heavy Tails*, Birkhauser, pp. 311–335.
- Koponen, I. (1995). Analytic approach to the problem of convergence of truncated Levy flights towards the Gaussian stochastic process, *Physical Review E* **52**: 1197–1199.
- Koutrouvelis, I. A. (1980). Regression-Type Estimation of the Parameters of Stable Laws, *Journal of the American Statistical Association* **75**: 918–928.
- Kuester, K., Mittnik, S., and Paolella, M.S. (2006). Value-at-Risk prediction: A comparison of alternative strategies, *Journal of Financial Econometrics* **4**(1): 53–89.
- Küchler, U., Neumann, K., Sørensen, M. and Streller, A. (1999). Stock returns and hyperbolic distributions, *Mathematical and Computer Modelling* **29**: 1–15.
- Leobacher, G. and Pillichshammer, F. (2002). A Method for Approximate Inversion of the Hyperbolic CDF, *Computing* **69**: 291–303.
- Lévy, P. (1925). *Calcul des Probabilités*, Gauthier Villars.
- Lombardi, M. J. (2007). Bayesian inference for α -stable distributions: A random walk MCMC approach, *Computational Statistics and Data Analysis* **51**(5): 2688–2700.
- Madan, D. B. and Seneta, E. (1990). The variance gamma (V.G.) model for share market returns, *Journal of Business* **63**: 511–524.
- Mandelbrot, B. B. (1963). The variation of certain speculative prices, *Journal of Business* **36**: 394–419.
- Mantegna, R. N. (1994). Fast, accurate algorithm for numerical simulation of Levy stable stochastic processes, *Physical Review E* **49**: 4677–4683.
- Mantegna, R. N. and Stanley, H. E. (1994). Stochastic processes with ultraslow convergence to a Gaussian: The truncated Lévy flight, *Physical Review Letters* **73**: 2946–2949.
- Marshall, A. W. and Olkin, I. (1988). Families of Multivariate Distributions, *Journal of the American Statistical Association* **83**: 834–841.
- Matacz, A. (2000). Financial Modeling and Option Theory with the Truncated Lévy Process, *International Journal of Theoretical and Applied Finance* **3**(1): 143–160.
- Matsui, M. and Takemura, A. (2006). Some improvements in numerical evaluation of symmetric stable density and its derivatives, *Communications in Statistics – Theory and Methods* **35**(1): 149–172.
- Matsui, M. and Takemura, A. (2008). Goodness-of-fit tests for symmetric stable distributions – empirical characteristic function approach, *TEST* **17**(3): 546–566.
- McCulloch, J. H. (1986). Simple consistent estimators of stable distribution parameters, *Communications in Statistics – Simulations* **15**: 1109–1136.
- McCulloch, J. H. (1997). Measuring Tail Thickness to Estimate the Stable Index α : A Critique, *Journal of Business & Economic Statistics* **15**: 74–81.
- McNeil, A. J., Rüdiger, F. and Embrechts, P. (2005). *Quantitative Risk Management*, Princeton University Press, Princeton, NJ.

- Michael, J. R., Schucany, W. R. and Haas, R. W. (1976). Generating Random Variates Using Transformations with Multiple Roots, *The American Statistician* **30**: 88–90.
- Mittnik, S., Doganoglu, T. and Chenyao, D. (1999). Computing the Probability Density Function of the Stable Pareto Distribution, *Mathematical and Computer Modelling* **29**: 235–240.
- Mittnik, S. and Paoletta, M. S. (1999). A simple estimator for the characteristic exponent of the stable Pareto distribution, *Mathematical and Computer Modelling* **29**: 161–176.
- Mittnik, S., Rachev, S. T., Doganoglu, T. and Chenyao, D. (1999). Maximum Likelihood Estimation of Stable Pareto Models, *Mathematical and Computer Modelling* **29**: 275–293.
- Nelder, J. A. and Mead, R. (1965). A Simplex Method for Function Minimization, *The Computer Journal* **7**: 308–313.
- Nelsen, R. B. (1999). *An Introduction to Copulas*, Springer, New York.
- Nolan, J. P. (1997). Numerical Calculation of Stable Densities and Distribution Functions, *Communications in Statistics – Stochastic Models* **13**: 759–774.
- Nolan, J. P. (1999). An Algorithm for Evaluating Stable Densities in Zolotarev's (M) Parametrization, *Mathematical and Computer Modelling* **29**: 229–233.
- Nolan, J. P. (2001). Maximum Likelihood Estimation and Diagnostics for Stable Distributions, in O. E. Barndorff-Nielsen, T. Mikosch, S. Resnick (eds.), *Lévy Processes*, Birkhäuser, Boston.
- Nolan, J. P. (2010). *Stable Distributions – Models for Heavy Tailed Data*, Birkhäuser, Boston. In progress, Chapter 1 online at academic2.american.edu/~jpnolan.
- Ojeda, D. (2001). *Comparison of stable estimators*, Ph.D. Thesis, Department of Mathematics and Statistics, American University.
- Paoletta, M. S. (2001). Testing the stable Pareto assumption, *Mathematical and Computer Modelling* **34**: 1095–1112.
- Paoletta, M. S. (2007). *Intermediate Probability: A Computational Approach*, Wiley, Chichester.
- Peters, G. W., Sisson, S. A. and Fan, Y. (2009). Likelihood-free Bayesian inference for α -stable models, *Preprint*: <http://arxiv.org/abs/0912.4729>.
- Poirot, J. and Tankov, P. (2006). Monte Carlo option pricing for tempered stable (CGMY) processes, *Asia-Pacific Financial Markets* **13**(4): 327–344.
- Prause, K. (1999). *The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures*, Ph.D. Thesis, Freiburg University, <http://www.freidok.uni-freiburg.de/volltexte/15>.
- Press, S. J. (1972). Estimation in Univariate and Multivariate Stable Distribution, *Journal of the American Statistical Association* **67**: 842–846.
- Press, W., Teukolsky, S., Vetterling, W. and Flannery, B. (1992). *Numerical Recipes in C*, Cambridge University Press. See also: <http://www.nr.com>.
- Pritsker, M. (2006). The hidden dangers of historical simulation, *Journal of Banking and Finance* **30**(2): 561–582.
- Protasov, R. S. (2004). EM-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ , *Statistics and Computing* **14**: 67–77.
- Rachev, S. and Mittnik, S. (2000). *Stable Pareto Models in Finance*, Wiley.
- Rosinski, J. (2007). Tempering stable processes, *Stochastic Processes and Their Applications* **117**(6): 677–707.
- Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes*, Chapman & Hall.
- Shuster, J. (1968). On the Inverse Gaussian Distribution Function, *Journal of the American Statistical Association* **63**: 1514–1516.
- Stahl, G. (1997). Three cheers, *Risk* **10**: 67–69.
- Stute, W., Manteiga, W. G. and Quindimil, M.P. (1993). Bootstrap Based Goodness-Of-Fit-Tests, *Metrika* **40**: 243–256.
- Trivedi, P.K., and Zimmer, D.M. (2005). Copula modeling: An introduction for practitioners, *Foundations and Trends in Econometrics* **1**(1): 1–111.
- Venter, J. H. and de Jongh, P. J. (2002). Risk estimation using the Normal Inverse Gaussian distribution, *The Journal of Risk* **4**: 1–23.

- Weron, R. (1996). On the Chambers-Mallows-Stuck Method for Simulating Skewed Stable Random Variables, *Statistics and Probability Letters* **28**: 165–171. *See also* R. Weron (1996) *Correction to: On the Chambers-Mallows-Stuck Method for Simulating Skewed Stable Random Variables*, Working Paper, Available at MPRA: <http://mpra.ub.uni-muenchen.de/20761/>.
- Weron, R. (2001). Levy–Stable Distributions Revisited: Tail Index > 2 Does Not Exclude the Levy–Stable Regime, *International Journal of Modern Physics C* **12**: 209–223.
- Weron, R. (2006). *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, Wiley, Chichester.
- Zolotarev, V. M. (1964). On representation of stable laws by integrals, *Selected Translations in Mathematical Statistics and Probability* **4**: 84–88.
- Zolotarev, V. M. (1986). *One–Dimensional Stable Distributions*, American Mathematical Society.

HSC Research Report Series 2010

For a complete list please visit <http://ideas.repec.org/s/wuu/wpaper.html>

- 01 *Models for Heavy-tailed Asset Returns* by Szymon Borak, Adam Misiorek and Rafał Weron
- 02 *FX Smile in the Heston Model* by Agnieszka Janek, Tino Kluge, Rafał Weron and Uwe Wystup
- 03 *Building Loss Models* by Krzysztof Burnecki, Joanna Janczura and Rafał Weron
- 04 *Ruin Probability in Finite Time* by Krzysztof Burnecki and Marek Teuerle
- 05 *Heavy-tailed distributions in VaR calculations* by Adam Misiorek and Rafał Weron